



Organisation de la chromatine et son lien avec la réplication de l'ADN

Benoît Moindrot

► To cite this version:

Benoît Moindrot. Organisation de la chromatine et son lien avec la réplication de l'ADN. Biologie cellulaire. Ecole normale supérieure de lyon - ENS LYON, 2012. Français. NNT : 2012ENSL0728 . tel-00733254

HAL Id: tel-00733254

<https://theses.hal.science/tel-00733254>

Submitted on 18 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

en vue de l'obtention du grade de

Docteur de l'École Normale Supérieure de Lyon – Université de Lyon

Discipline : Sciences de la vie

Préparée au Laboratoire Joliot Curie (USR3010)

École doctorale : Biologie Moléculaire et Cellulaire, Intégrative

Présentée et soutenue publiquement le 11 Juillet 2012

par **Benoît MOINDROT**

Organisation de la chromatine et son lien avec la réplication de l'ADN

Directeur de thèse : Philippe BOUVET

Après l'avis de : Jean-Marc LEMAÎTRE
Peter MEISTER

Devant la commission d'examen formée de :

Philippe	BOUVET	Directeur
Edith	HEARD	Membre
Saadi	KHOCHBIN	Membre
Jean-Marc	LEMAÎTRE	Rapporteur
Peter	MEISTER	Rapporteur
Fabien	MONGELARD	Membre

Résumé

L'organisation de la chromatine a une importance fonctionnelle pour contrôler le programme d'expression des gènes. Par contre, les liens qui l'unissent au déroulement de la réplication de l'ADN sont beaucoup moins connus.

Grâce à des approches basées sur la capture d'interactions chromosomiques et sur l'imagerie cellulaire, nous avons étudié les liens entre le repliement à grande échelle de la chromatine et le timing de réplication. Cette analyse, effectuée dans des cellules humaines lymphoblastoïdes, des cellules mononucléées du sang (PBMC) et des cellules issues d'une leucémie myéloïde à caractère érythrocytaire, a permis l'identification de domaines structuraux du noyau. Ces domaines sont relativement isolés les uns des autres et leurs frontières coïncident avec les zones d'initiation précoce. De plus, notre étude montre que ces zones d'initiation précoce interagissent préférentiellement, aussi bien entre voisins immédiats (séparation génomique de l'ordre de la mégabase) que le long du chromosome entier. Les loci répliqués tardivement interagissent eux-aussi avec leurs homologues, conduisant, dans l'espace nucléaire, à une ségrégation des loci en fonction de leur timing de réplication. Ces résultats sont soutenus par des mesures de distances sur des hybridations *in-situ* qui montrent que les loci répliqués en début de phase S sont plus proches qu'attendus.

Nos travaux révèlent enfin que l'organisation de la chromatine est similaire dans des cellules en phase G0 (PBMC dormantes), démontrant qu'elle n'est pas spécifique des cellules en phase S.

Pris ensemble, ces résultats apportent des preuves directes d'une organisation robuste de la chromatine, partagée par les cellules en cycle et dormantes, et corrélée au timing de réplication à différentes échelles.

Mot-clés : Chromatine, Réplication de l'ADN, Noyau cellulaire, Capture de conformation chromosomique, Génome

Abstract

Chromatin organization is of functional significance to control the gene expression program. However, its interplay with DNA replication program is less known.

Though the capture of chromosomal interactions and cell imaging, we studied the links between the high-order folding of chromatin and the replication timing. This analysis, which has been performed in human lymphoblastoid cells, in peripheral blood mononuclear cells (PBMC), and in a myeloid leukemia cell line with erythroid properties, allowed the identification of structural domains in the nucleus. These domains are quite isolated from each other and their boundaries coincide with early-initiation zones. In addition, our study shows that these early-initiation zones preferentially interact with each other. These interactions have been observed between neighboring early-initiation zones (genomic separation around 1 to few megabases) but also along the whole chromosome. The late-replicated loci interact with their counterparts as well, leading to a nuclear segregation of the loci according to their replication timing. These results are sustained by distance measurements in *in-situ* hybridizations which show that loci replicated at the beginning of S-phase are closer than expected.

Our work also reveals that chromatin organization is similar in cells blocked in G0 phase (quiescent PBMC), demonstrating that it does not result from the cells in S-phase.

Taken together, these results provide direct clues for a robust chromatin organization, common to cycling and resting cells, and related to the replication timing at different scales

Keywords : Chromatin, DNA Replication, Cell nucleus, Chromosome Conformation Capture, Genome

Remerciements

J'aimerais vivement remercier le Dr. Edith Heard, le Dr. Jean-Marc Lemaître, le Dr. Peter Meister et le Dr. Saadi Khochbin pour avoir jugé mon travail de thèse. Merci également pour l'enrichissante discussion lors de la phase des questions : certains points que vous avez soulevés confèrent à mes travaux des perspectives que je n'avais pas imaginées.

Je souhaite également remercier le Pr. Philippe Bouvet qui m'a permis de faire cette thèse dans son équipe sur cette thématique particulièrement distante du nucléole et de la nucléoline. Merci d'avoir permis que ce travail voie le jour, d'y avoir apporté ton précieux avis.

Je souhaite particulièrement remercier le Dr. Fabien Mongelard pour m'avoir encadré au cours de ces 4 dernières années. Sans ton idée folle en 2008, il n'y aurait jamais eu de "Chromosome Conformation Capture" au labo. Merci pour nos discussions au jour le jour, pour ton soutien sans faille, et pour ta légendaire patience (il en a fallu, non ?). A plusieurs reprises, ces 4 dernières années n'auront pas été faciles, mais tu as toujours été là. J'ai eu de la chance d'avoir un encadrant comme toi.

Je pense également à Alain Arneodo, Benjamin Audit, Cédric Vaillant, Antoine Baker, Hanna Julienne, Claude Thermes, et Olivier Hyrien : Merci pour le timing de répllication et les N-domaines, merci pour nos discussions fructueuses (réunions ANR ou autres), pour les conseils notamment en programmation, et j'en oublie certainement.

Je souhaite également vivement remercier Wouter de Laat qui m'a accueilli dans son laboratoire pour apprendre le 4C, et Petra Klous pour m'en avoir montré tous les trucs/astuces/subtilités qui ne sont jamais écrites dans les protocoles. Merci également à David Gilbert pour nos enrichissantes discussions.

Je souhaite aussi remercier d'autres collègues. Merci à Karine pour ses conseils en imagerie cellulaire et à Kévin pour son efficacité en assistance informatique. Merci à Sadhan et Rong pour nos discussions plus ou moins scientifiques, nos plaisanteries à la paillasse ou en salle culture ; Merci à Xavier pour avoir en plus ajouté beaucoup de bonne humeur dans le bureau (*maintenant, tu es le gardien du temple et de la macro ImageJ : j'ai amplement confiance*) ; Merci à Zophia (*plus jamais je ne te piquerai ta p2*), Elodie (*Génial tes propositions d'épigraphe*), Hélène (*Merci pour ton super polo*), Maggy (*la meilleure en biologie de la tomate!*) pour les petites pauses improvisées, pour l'horoscope, pour votre soutien, pour votre écoute et pour votre rôle majeur afin que le laboratoire continue de tourner.

J'ai également une pensée émue pour les thésards en pleine rédaction (Salim, Cristina, *Courage, vous touchez au but*), et celui qui va bientôt commencer (Sadhan *Good luck and all the best for the next few months*).

Merci aussi aux fidèles du Ninkasi qui n'avaient pas encore été cités pour tous les bons moments que nous avons passés ensemble (Arnaud, Elise, Johan, Marc, Maryline, Monica, Pascale).

Parce que tout ne se limite pas au laboratoire. . .

Merci à Catherine et Xavier pour m'avoir initié (dans le désordre) à Rock-band, la bière, les belles photos, les règles du basket (oui, j'en suis resté là). . . ; à Juliette et Rémi pour l'initiation à l'humour "machette / Chuck Norris", à Radio-head, au domptage de panthère. . . ; à Angèle pour la découverte du Jura avec son précieux terroir, ainsi que pour avoir décroché la Lanterne après un passage par l'Académie. . . ; à Nico pour ton entrain, pour contrôler l'enfant terrible sur la banquette arrière, pour être le Lucky Luke de l'endormissement (qu'est-ce qu'on a pu en rire). . . Merci à tous les 6 d'être présent le temps des vacances, d'un weekend, d'une soirée, d'un coup de téléphone. Votre amitié est extrêmement précieuse, elle m'a sans conteste porté pendant ces 4 ans.

Un énorme merci également à mes parents qui m'ont toujours soutenu, encouragé, aidé, bref, merci d'avoir toujours été là pour moi ; merci aussi à mon frerot qui m'a souvent conseillé (entre autre pour les détente cinéma). Merci au reste de ma famille pour leur soutien indéfectible.

Enfin, mes derniers remerciements sont pour Muriel : merci d'avoir relu mon manuscrit, d'avoir toujours cru en moi, d'avoir partagé et raisonné mes doutes, merci aussi pour ton aide dans les moments plus compliqués, merci aussi pour ton attention et tous ces petits instants de notre vie à deux qui font mon bonheur. Merci du fond du cœur.

TABLE DES MATIÈRES

PARTIE A INTRODUCTION

1	DOMAINES NUCLÉAIRES ET CHROMATINE	PAGE 23
1.1	Organisation structurale et fonctionnelle de la chromatine	24
	Le nucléosome, l'unité structurale de la chromatine — 24	
	Un nombre limité d'états chromatiniens différents — 28	
	De la fibre de 10 nm à la fibre de 30 nm — 28	
	De nombreux indices pour l'existence de boucles géantes de chromatine — 30	
	Territoires chromosomiques — 33	
1.2	Des repliements contrôlés et dynamiques	37
	Méthodes de la famille 3C — 37	
	Des boucles de chromatine locales en lien avec leurs fonctions — 42	
	Repliement et définition de domaines chromatiniens — 44	
	Repliement et établissement de contacts entre chromosomes — 48	
1.3	Structure nucléaire et localisation des contacts	52
	Structure du noyau — 52	
	Localisation à un compartiment actif — 53	
	Localisation à un compartiment répressif — 56	
2	LA RÉPLICATION DE L'ADN	PAGE 59
2.1	Mécanismes généraux de la réplication de l'ADN	60
	Déroutement de la réplication — 60	
	Les origines de réplication identifiées chez l'homme par biologie moléculaire — 63	
	Contrôle du choix des ORI et de leur activation coordonnée — 64	
2.2	Approches génomiques et à haut-débit pour l'étude de la réplication	66
	Biais de composition en nucléotides et réplication — 67	
	Identification de sites de liaison du complexe pré-réplcatif (pre-RC) — 69	
	Détection de sites d'initiation — 70	
	Établissement du timing de réplication — 71	
2.3	Structuration de la chromatine associée à la réplication	78
	Réplication et boucles d'ADN — 79	
	Les foyers de réplication — 82	
	Contacts chromatiniens et réplication — 86	
	PROJET DE THÈSE	PAGE 91

PARTIE B RÉSULTATS

3	MISE EN PLACE D'UNE STRATÉGIE 4C-SEQ	PAGE 95
3.1	Choix de la région et du type cellulaire.	97
	Sélection d'une région à timing conservé et contrasté (chr5 :150-170Mb) — 97	
	Choix du type cellulaire — 99	
3.2	De la librairie 4C au séquençage haut-débit	100
	Préparation des librairies 4C — 100	
	Positionnement et conception des oligonucléotides 4C — 104	
	Profils des PCR 4C — 104	
	Des PCR 4C-Seq au séquençage haut-débit — 107	
3.3	Bioinformatique et traitement du séquençage haut-débit	110
	Dé-multiplexer les lectures 4C — 110	
	Alignement sur le génome — 112	
	Filtrage des lectures en fonction de leur alignement — 115	

4 CORRESPONDANCE ENTRE TIMING DE RÉPLICATION ET ORGANISATION DE LA CHROMATINE PAGE 117

4.1	Introduction121
4.2	Résultats122
	Les zones d'initiation précoce délimitent des domaines d'auto-interaction —	128
	Les interactions à longue-distance reproduisent le timing de réplication —	132
	Des mesures de distances confirment la proximité des loci précoces —	138
4.3	Discussion140
4.4	Méthodes143

5 TOPOLOGIE DE LA CHROMATINE : EXTENSION DE L'ANALYSE À D'AUTRES ÉCHELLES, POINTS DE VUE ET TYPES CELLULAIRES PAGE 147

5.1	Conformation de la chromatine dans des cellules pro-érythrocytaires.148
	Conformation de la chromatine à proximité des points de vue —	148
	Partenaires préférentiels à longue-distance —	151
5.2	Organisation locale de la chromatine à timing constant152
	Détail de la région et PCR 4C-Seq —	152
	Analyse des profils 4C —	153
5.3	Analyse des partenaires chromatiniens d'un plateau précoce156
5.4	Les loci précoces sur plusieurs Mb <i>trans</i> -interagissent158
	Matériel et méthodes : Analyse des <i>trans</i> -interactions dans le Hi-C —	158
	Proportion d'interactions en <i>trans</i> le long du chromosome —	159
	Chromosomes préférentiellement contactés —	164

PARTIE C DISCUSSION

6 DISCUSSION ET CONCLUSION PAGE 169

6.1	L'organisation de la chromatine à grande échelle est corrélée au timing de réplication à différentes échelles.169
	Conformation à l'échelle de quelques mégabases —	169
	Conformation à l'échelle de plusieurs dizaines de Mb —	173
	Conformation à l'échelle du territoire chromosomique —	173
6.2	Les profils d'interactions sont mieux décrits par le timing que par le niveau de transcription174
6.3	Les contacts chromatiniens dans les cellules en phase G0177
6.4	Forces pour établir cette organisation de la chromatine179
6.5	Conclusion et perspectives180

BIBLIOGRAPHIE PAGE 182

LISTE DES FIGURES

Figure 1–	Existence d’une unité structurale de la chromatine : le nucléosome.25
Figure 2–	Domaine "Histone Fold"25
Figure 3–	Structure et dimensions du nucléosome26
Figure 4–	Modifications post-traductionnelles des histones de cœur27
Figure 5–	Principaux modèles pour l’organisation de la fibre de 30 nm29
Figure 6–	Chromosome en métaphase dépourvu d’histones observé par microscopie électronique	31
Figure 7–	Repliement de la chromatine – modèles "polymères" avec boucles33
Figure 8–	Existence de territoires Chromosomiques34
Figure 9–	Repliement hiérarchique de la chromatine36
Figure 10–	Aperçu des différentes méthodes de la famille 3C38
Figure 11–	Contacts capturés par 3C entre des enhancers et leurs gènes cibles42
Figure 12–	Contact entre les régions différemment méthylées du locus H19/Igf243
Figure 13–	Existence de domaines chromatinien45
Figure 14–	Structure du noyau53
Figure 15–	Localisation de loci à différents sous-compartiments nucléaires55
Figure 16–	La réplication de l’ADN chez les eucaryotes61
Figure 17–	Quelques points de régulation de la réplication65
Figure 18–	Détection de N-domaines de timing d’après les biais de composition en nucléotides . .	.68
Figure 19–	Différentes méthodes permettant la détection d’origines de réplication69
Figure 20–	Propriétés des régions précoces et tardives dans les cellules humaines K56273
Figure 21–	Calcul du timing moyen de réplication76
Figure 22–	Foyers de réplication au cours de la phase S83
Figure 23–	Coordination de deux fourches de réplication au sein du même foyer de réplication .	.87
Figure 24–	Hi-C et réplication – les limites89
Figure 25–	Choix d’une région à timing de réplication conservé96
Figure 26–	Timing de réplication (MRT) de la région chr5 :150–170 Mb98
Figure 27–	Contrôle de la digestion sur chromatine cross-linkée102
Figure 28–	Contrôle de la ligation et de la ligation de la digestion par Hin1II103
Figure 29–	Positionnement des oligonucléotides 4C105
Figure 30–	Contrôle du bon déroulement des PCR 4C106
Figure 31–	10 réplicats de la même PCR 4C-Seq108
Figure 32–	Possibilité de multiplexer les PCR 4C avant séquençage109
Figure 33–	Des lectures aux profils 4C - les étapes bioinformatiques111
Figure 34–	Alignement sur le génome humain113
Figure 35–	Alignement sur le génome - distance au site HindIII le plus proche115

Figure 36–	Local chromatin properties around the ten selected viewpoints in PBMC124
Figure 37–	4C interaction profiles associated with replication timing in lymphoblastoid cells128
Figure 38–	Entry of PHA-stimulated cells into cell cycle130
Figure 39–	4C interaction profiles associated with replication timing in PBMC and PHA-stimulated PBMC131
Figure 40–	The detection of long-range interactions is independent on the variable choice132
Figure 41–	Chromosomal domains with opposite replication timing segregate in the nuclear space	.134
Figure 42–	LRIF signals in PHA-stimulated PBMC136
Figure 43–	Viewpoints localized on local timing peaks mostly interact with other timing-peaked regions136
Figure 44–	Inter-chromosomal LRIF signals in resting PBMC137
Figure 45–	FISH distance measurements confirm the physical proximity between early replicated regions139
Figure 46–	Hi-C interaction profile across timing peaks140
Figure 47–	Replication domains correspond to structural self-interacting domains141
Figure 48–	Profils d'interactions locales des cellules lymphoblastoïdes et K562149
Figure 49–	Profils d'interactions à longue-distance des cellules lymphoblastoïdes et K562150
Figure 50–	PCR 4C avec comme point de vue l'origine de réplication à proximité de c-myc153
Figure 51–	Conformation de la chromatine au sein d'un plateau précoce154
Figure 52–	Un plateau précoce interagit préférentiellement avec d'autres loci précoces156
Figure 53–	Proportion d'interactions en <i>trans</i> : variation d'un chromosome à l'autre160
Figure 54–	<i>Trans</i> -partenaires le long des chromosomes 1 et 12162
Figure 55–	Organisation topologique de la chromatine en suivant le timing de réplication172
Figure 56–	Le taux de transcription n'explique pas à lui-seul les contacts chromatiniens174

LISTE DES TABLEAUX

Tableau 1–	Read number and distribution123
Tableau 2–	Proportion of reads per chromosome123
Tableau 3–	Primers used for the 4C-Seq experiment144
Tableau 4–	BAC clones used for the FISH experiments145

LISTE DES ABRÉVIATIONS

Sigle	Signification
3C	Capture de Conformation Chromosomique (Chromosome Conformation Capture)
3D	Tridimensionnel
4C	Capture circulaire de conformation chromosomique
4C-Seq	Capture circulaire de conformation chromosomique couplée à un séquençage haut-débit
5C	Copie-carbone de Capture de conformation chromosomique
AA	Acide aminé
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
ARNm	ARN messenger
ARS	Séquence de réplication autonome (Autonomously Replicating Sequence)
ATP	Adénosine-5'-triphosphate
BAC	Chromosome artificiel bactérien
bp	Paire de base (acide nucléique)
BrdU	Bromodéoxyuridine
ChIP	Immunoprécipitation de chromatine
CHO	Cellules d'Ovaire de Hamster Chinois (Chinese Hamster Ovary)
chr	Chromosome
CT	Territoire chromosomique (Chromosome Territory)
DNase	Désoxyribonucléase
ENCODE	Encyclopédie d'éléments ADN (ENCyclopedia Of DNA Elements)
FACS	Tri cellulaire en fonction du niveau de fluorescence (Fluorescence-activated cell sorting)
FISH	Hybridation <i>in-situ</i> fluorescente
GFP	Protéine fluorescente verte (Green Fluorescent Protein)
HAT	Histone acétyltransférase
HDAC	Histone déacétylase
ICR	Région de contrôle par l'empreinte parentale (Imprinting Control Region)
kb	Kilobase (= 1000 bp)
LCR	Région de contrôle du locus (Locus Control Region)
LRIF	Fréquence d'interaction à longue-distance (Long-Range Interaction Frequency)
MAR	Région d'ancrage à la matrice nucléaire (Matrix Attachment Region)
Mb	Mégabase (= 1000 kb)
Mcm	Maintenance du mini chromosome (Mini Chromosome Maintenance)
MRT	Timing de réplication moyen (Mean Replication Timing)
ORC	Complexe de reconnaissance des origines (Origin Recognition Complex)
ORI	Origine de réplication

Sigle	Signification
PBMC	Cellules mononucléées du sang périphérique
PCNA	Antigène nucléaire de cellules proliférantes (Proliferating cell nuclear antigen)
PCR	Réaction de polymérisation en chaîne (Polymerase Chain Reaction)
PHA	Phytohémagglutinine
PRE	Elements de réponse aux complexes Polycomb (Polycomb Response Elements)
pre-RC	Complexe pré-répliatif
Ridge	Région d'expression génique forte (Regions of Increased Gene Expression)
RNase	Ribonucléase
RPA	Protéine de Réplication A (Replication Protein A)
RWGL	Marche aléatoire avec boucles géantes (Random-walk giant Loop)
S/MAR	Région d'ancrage à la matrice / au squelette (Scaffold / Matrix Attachment Region)
TCC	Capture de conformation ancrée (Tethered conformation capture)

Partie A

Introduction

A

L'acide désoxyribonucléique (ADN), présent dans chaque cellule, constitue le support de l'information génétique. Pour que chaque cellule possède le même patrimoine, l'ADN est dupliqué à l'identique avant que la cellule ne se divise. Cette duplication (aussi appelée réplication), dure en moyenne 8 h et permet aux deux cellules filles de recevoir une copie intégrale de l'information génétique.

Chaque cellule contient approximativement 2 mètres d'ADN embobinés dans un noyau dont le diamètre est 200 000 fois plus petit. Pour être opérationnel par la cellule, l'ADN n'est pas entassé de manière aléatoire dans le noyau. Prenons un seul exemple. Au cours du développement embryonnaire, certains gènes sont activés et d'autres sont réprimés. C'est le cas du locus *Dppa2* chez la souris qui est exprimé dans les cellules souches, mais pas dans les cellules différenciées. Ce changement d'expression s'accompagne d'un changement dans le positionnement dans le noyau : dans les cellules souches, le locus *Dppa2* est localisé au centre du noyau, alors qu'il adopte une position beaucoup plus périphérique dans les cellules différenciées [Hiratani et al., 2010]. En parallèle, ce locus, qui faisait partie des premières séquences dupliquées dans les cellules souches, est répliqué beaucoup plus tardivement dans les cellules différenciées. Pour résumer, au cours du développement, le locus *Dppa2* est d'abord exprimé / dupliqué précocement / localisé au centre du noyau, puis il devient réprimé / dupliqué tardivement / localisé en périphérie du noyau. Cet exemple n'est pas un cas isolé car de nombreux loci suivent la même tendance.

Si aucun lien de causalité ne peut être avancé, cette observation atteste de l'existence d'une relation entre l'organisation nucléaire d'un côté (comment l'ADN se replie), et de l'autre l'expression et la réplication. Ce constat soulève plusieurs questions. Comment le noyau est-il organisé, et quelles méthodologies peuvent être utilisées pour aborder cette question ? Quels sont ces liens entre organisation nucléaire et expression / réplication ? A quel point sont-ils robustes ?

Ces questions constitueront le fil conducteur de ce manuscrit. Dans la partie résultat, nous nous focaliserons sur les liens entre organisation nucléaire et réplication. Néanmoins, l'introduction sera plus générale et définira le cadre de cette étude. Nous aborderons donc également les liens entre organisation nucléaire et la transcription qui sont, de surcroît, souvent plus étudiés. Au cours de cette introduction, nous suivrons donc la démarche suivante :

- Tout d'abord nous détaillerons les preuves démontrant que le noyau est un organe organisé. Ceci nous amènera à étudier comment l'ADN se replie, quelles sont les preuves que ce repliement n'est pas aléatoire, et quelles sont les structures nucléaires qui stabilisent ces repliements. Au cours de ce chapitre, l'accent sera volontairement porté sur les liens entre organisation nucléaire et transcription car ils sont mieux connus (ceux impliquant la réplication seront décrits dans le chapitre suivant).
- Puis nous aborderons comment se déroule la réplication, tout d'abord en décrivant succinctement les acteurs enzymatiques impliqués, puis en détaillant les méthodes permettant son étude sur le génome entier. En fin de chapitre, nous discuterons les liens entre l'organisation nucléaire et le déroulement de la réplication avant de conclure sur les questions encore en suspens.

Domaines nucléaires et chromatine

1.1	Organisation structurale et fonctionnelle de la chromatine	24
1.1.1	Le nucléosome, l'unité structurale de la chromatine	24
	<i>Cœur du nucléosome et enroulement d'ADN</i>	24
	<i>Modifications post-traductionnelles des histones</i>	26
	<i>L'hypothèse d'un code "histone"</i>	28
1.1.2	Un nombre limité d'états chromatiniens différents.	28
1.1.3	De la fibre de 10 nm à la fibre de 30 nm	28
1.1.4	De nombreux indices pour l'existence de boucles géantes de chromatine	30
	<i>Les chromosomes en écouvillon</i>	30
	<i>Les chromosomes dépourvus d'histones</i>	30
	<i>La matrice nucléaire</i>	31
	<i>Marche de polymères d'après des hybridations in-situ</i>	32
1.1.5	Territoires chromosomiques	33
	<i>Découverte des territoires chromosomiques par micro-irradiation UV</i>	34
	<i>Chromosome Painting</i>	35
	<i>Arrangement non-aléatoire des territoires chromosomiques</i>	35
	<i>Imbrication des territoires chromosomiques</i>	35
1.2	Des repliements contrôlés et dynamiques	37
1.2.1	Méthodes de la famille 3C	37
	<i>Aperçu des différentes méthodes de la famille 3C</i>	37
	<i>Discussion des champs d'application et limites du 4C et du Hi-C</i>	40
1.2.2	Des boucles de chromatine locales en lien avec leurs fonctions.	42
	<i>Contacts enhanceur – gène</i>	42
	<i>Contacts silenceur – gène</i>	43
	<i>Contacts entre régions régulatrices</i>	43
1.2.3	Repliement et définition de domaines chromatiniens.	44
	<i>Segmentation en domaines fonctionnels</i>	44
	<i>Segmentation en domaines structuraux</i>	46
	<i>Maintien des frontières par CTCF et les cohésines</i>	47
1.2.4	Repliement et établissement de contacts entre chromosomes	48
	<i>Éviction du territoire chromosomique</i>	49
	<i>Formation de trans-contacts</i>	49
	<i>Approches plus globales</i>	50
1.3	Structure nucléaire et localisation des contacts	52
1.3.1	Structure du noyau	52
1.3.2	Localisation à un compartiment actif	53
	<i>Les usines de transcription</i>	53
	<i>Les corps d'épissage ou speckles</i>	53
	<i>Les corps PML</i>	54
1.3.3	Localisation à un compartiment répressif	56
	<i>La lamine nucléaire</i>	56
	<i>Les corps Polycomb</i>	56
	<i>Périphérie du nucléole</i>	57

L'ADN n'est pas nu dans la cellule. Il est accompagné de protéines, formant ce qu'on appelle la chromatine (du grec *chromos*, couleur, en référence aux travaux de W. Flemming qui remarqua des structures dans le noyau cellulaire absorbant fortement l'aniline, un colorant basique). L'organisation de la chromatine dans le noyau peut être décrite à différentes échelles qui seront abordées dans ce chapitre. Nous insisterons particulièrement sur les preuves que le repliement de la chromatine n'est pas aléatoire, en suivant le cheminement suivant :

- 1 nous verrons que le repliement de la chromatine s'effectue de manière hiérarchique ;
- 2 nous démontrerons que ces repliements, qui sont contrôlés et non aléatoires, impliquent la formation de contacts préférentiels entre certaines portions de chromatine ;
- 3 et enfin, nous considérerons le noyau dans son ensemble et verrons que les contacts préférentiels peuvent s'établir au sein de sous-compartiments.

L'implication de la réplication sera volontairement mise de côté dans ce premier chapitre. Je vous propose en effet d'étudier dans un premier temps l'organisation de la chromatine au sein du noyau dans son ensemble, avant d'insister spécifiquement, dans le chapitre 2, sur les liens entre agencement de la chromatine et réplication.

1.1- Organisation structurale et fonctionnelle de la chromatine

1.1.1- Le nucléosome, l'unité structurale de la chromatine

La digestion partielle de l'ADN par des endonucléases génère des fragments d'ADN ayant une taille multiple de 200 bp. Ces fragments peuvent aisément être visualisés sur un gel de polyacrylamide (figure 1A) et ne sont pas obtenus lorsque la même expérience est effectuée sur de l'ADN purifié [Hewish and Burgoyne, 1973]. Ceci a permis de proposer qu'il existe, au sein du noyau, une unité structurale de base. Cette interprétation a été largement confortée par des observations de microscopie électronique (figure 1B) montrant un chapelet de particules reliées les unes aux autres par un filament d'ADN [Olins and Olins, 1974]. Cette particule élémentaire a été nommée "nucléosome" [Oudet et al., 1975].

a- Cœur du nucléosome et enroulement d'ADN

Le nucléosome est formé de 8 histones de cœur ($[\text{H2A}, \text{H2B}, \text{H3 et H4}] \times 2$) autour desquelles s'enroule 146 bp d'ADN. Les histones de cœur forment une famille de petites protéines basiques très conservées au cours de l'évolution. Elles sont constituées d'une partie N-terminale non-structurée et d'un domaine structuré (motif "histone fold"). La structure précise d'un nucléosome a été déterminée par diffraction aux rayons X [Luger et al., 1997]. Le motif "histone fold", d'environ 70 résidus, est composé de trois hélices alpha connectées par de courtes boucles (figure 2).

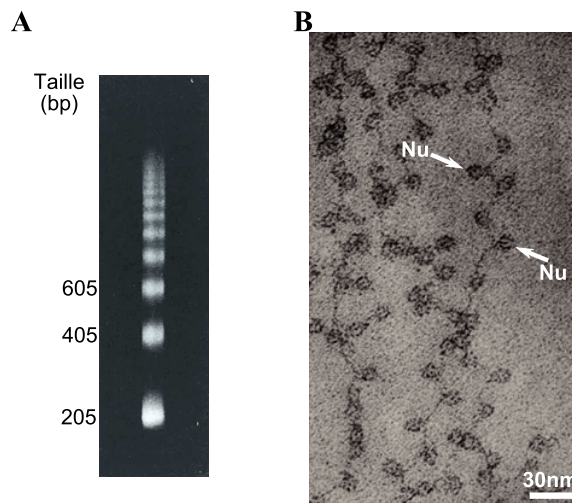


FIGURE 1 – Existence d’une unité structurale de la chromatine : le nucléosome. **A-** Digestion ménagée de chromatine à la nucléase micrococcale (d’après le chapitre 29 de [Lewin \[2008\]](#)). **B-** Chromatine isolée en faible condition ionique et observée par microscopie électronique à transmission. Adapté avec autorisation de Macmillan Publishers Ltd : Nature Reviews Molecular Cell Biology Olins, D. E. and Olins, A. L. (2003). Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol*, 4(10):809–814. Copyright 2003

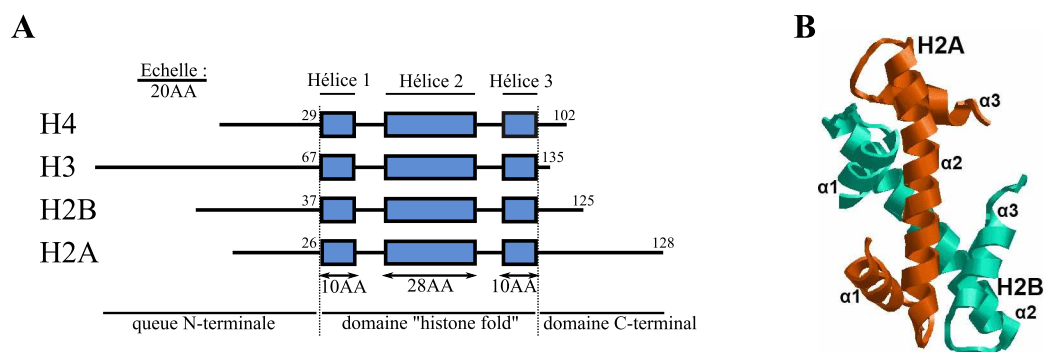


FIGURE 2 – Le domaine "histone fold" chez l’homme. **A-** Alignement des 4 histones de cœur, les hélices alpha sont représentées par les rectangles bleus. De part et autre du motif "histone fold" central se trouve les queues N-terminales et les domaines C-terminaux. AA = acide aminé. **B-** Structure tridimensionnelle des domaines "histone fold" de H2A et H2B qui s’emboîtent en poignée de main. Les trois hélices alpha de chaque histone sont notées $\alpha 1$, $\alpha 2$ et $\alpha 3$. Structure obtenue à partir du fichier 1AOI.pdb.

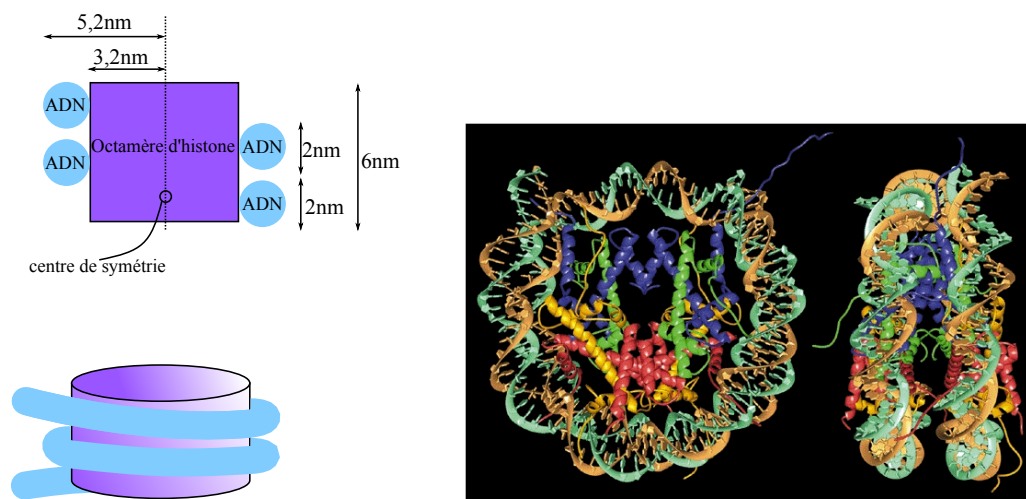


FIGURE 3 – Structure et dimensions du nucléosome. La figure de droite correspond à la structure tridimensionnelle du nucléosome obtenue par [Luger et al. \[1997\]](#) ; les croquis de gauche schématisent la structure du nucléosome ainsi que ses principales dimensions. La figure de droite est adaptée avec autorisation de Macmillan Publishers Ltd : Nature Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648):251–260. Copyright 1997.

Les 8 histones de cœur forment ensemble un cylindre plat de 6.4 nm de diamètre pour 6 nm d'épaisseur. Autour de ce cylindre s'enroule 1.7 tour d'ADN (soit 146 bp) en une hélice gauche (figure 3).

b- Extrémités N-terminale et C-terminale des histones et modifications post-traductionnelles

De part et d'autre du domaine "histone fold", les histones possèdent des extrémités N- et C-terminales (figure 2A). Ces extrémités sont nettement moins structurées que les domaines de cœur et se retrouvent en périphérie du disque nucléosomal (bien qu'invisibles car non cristallisées, des portions d'extrémités N-terminales sont représentées sur la figure 3). Les extrémités N-terminales ne participent pas directement à la structure du nucléosome. Elles interviennent plutôt dans des interactions avec les nucléosomes voisins, et participent donc, de proche en proche, à une compaction de niveau supérieur de la chromatine que l'on détaillera dans les parties suivantes.

Les histones peuvent être modifiées chimiquement. Ces modifications chimiques, qui sont mises en place après la synthèse protéique des histones (modifications post-traductionnelles) sont majoritairement retrouvées au niveau des queues N-terminales des histones (figure 4).

Acétylation des histones Ajouté sur des lysines, le groupement acétyle neutralise partiellement les charges positives des histones, et contribue donc à diminuer leur affinité pour l'ADN. L'acétylation des histones est effectuée par des acétyltransférases (HAT) tandis que la suppression du groupement acétyle est effectuée par des histones déacétylases (HDAC). Les 4 histones de cœur peuvent être acétylées (figure 4), et les régions du génome hyper-acétylées sont principalement retrouvées au niveau des gènes exprimés ou des séquences régulatrices [[Li et al., 2007](#)]. En plus de son effet sur la transcription, le niveau d'acétylation intervient

1.1- Organisation structurale et fonctionnelle de la chromatine

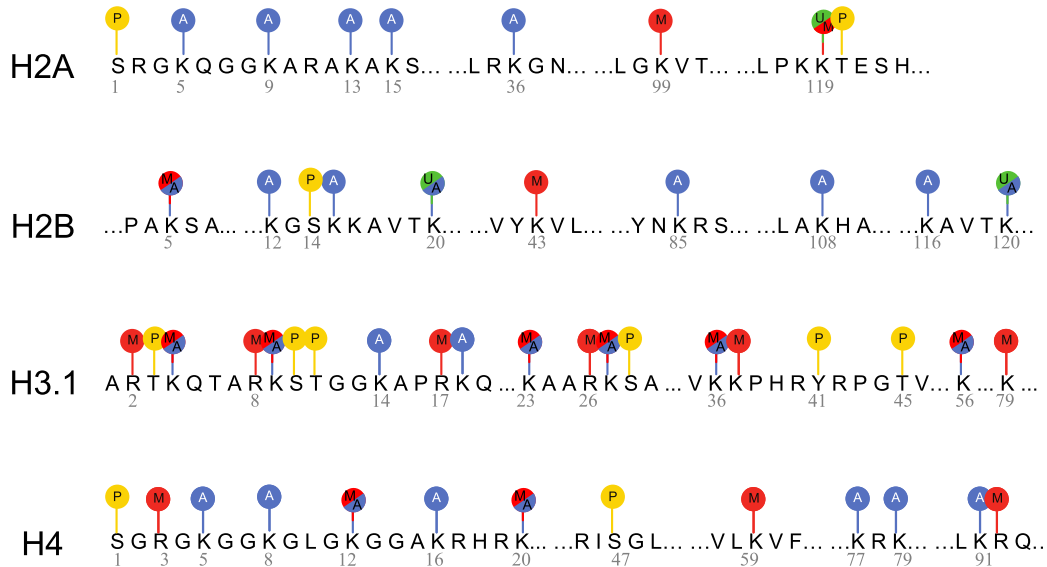


FIGURE 4 – Les principales modifications post-traductionnelles des histones de cœur. Les SUMOylation, ADP-ribosylation, biotinylation, hydroxylation et crotonylation ne sont pas représentées. D'après [Portela and Esteller \[2010\]](#)

également au niveau de la réparation [[Ikura et al., 2000](#); [Bird et al., 2002](#)] et de la réplication [[Miotto and Struhl, 2010](#)] de l'ADN.

Méthylation des histones Un, deux ou trois groupements méthyles peuvent être ajoutés sur les fonctions amines des lysines et des arginines. Les histones méthyltransférases sont impliquées dans l'ajout de ces groupements méthyles, alors que les histones déméthylases les enlèvent.

La position et le nombre de groupement méthyles va avoir des effets différents sur la chromatine [[Izzo and Schneider, 2010](#)].

- Certaines méthylations sont associées à une chromatine active, par exemple :
 - o sur des lysines : H3K4me(1/2/3), H3K36me(1/2/3), H4K20me1 ;
 - o sur des arginines : H3R17me(1/2).
- D'autres méthylation sont associées à une chromatine répressive, par exemple :
 - o sur des lysines : H3K9me(1/2/3), H3K27me(2/3), H4K20me3
 - o sur des arginines : H3R2me2.

Autres modifications des histones D'autres modifications d'histones existent :

- Les sérines, tyrosines et thréonines peuvent être phosphorylées (figure 4). Par exemple, le niveau global de sérine 10 phosphorylée sur l'histone H3 augmente de façon drastique en fin de phase G2, persiste en prophase et métaphase, puis diminue à la sortie de la phase M. Néanmoins, certains gènes actifs restent phosphorylés en interphase [[Prigent and Dimitrov, 2003](#)].
- H2BK120 peut être ubiquitinée et il semblerait qu'il existe un lien étroit entre l'ubiquitination de cette lysine et la méthylation de H3K4 [[Kim et al., 2009](#)]. Différentes modifications d'histones sont donc couplées pour contrôler l'expression des gènes.
- Certains acides aminés des histones peuvent également porter une SUMOylation, une ADP-ribosylation, une biotinylation, une hydroxylation, ou une crotonylation.

c- L'hypothèse d'un code "histone"

Strahl and Allis [2000] ont proposé que ces modifications post-traductionnelles des queues des histones pouvaient constituer un "code" compréhensible par la cellule. Ceci suppose l'existence de protéines capables d'établir le code ("writer"), de lire ce code ("reader"), et de l'effacer au besoin ("eraser"). Pour la plupart des modifications post-traductionnelles d'histones, des complexes protéiques sont connus dans les 3 catégories [Kouzarides, 2007a,b; Allis et al., 2007; Yun et al., 2011].

L'enroulement de l'ADN autour des 8 histones de cœur constitue un premier niveau de repliement. Le nucléosome correspond ainsi à une unité structurale de base de la chromatine. Les histones constituant ce nucléosome peuvent être modifiées chimiquement, ce qui aura des effets sur la transcription des gènes.

1.1.2- Un nombre limité d'états chromatiniens différents

Avec la très grande diversité des modifications post-traductionnelles d'histones et la diversité des complexes responsables de ces modifications, on pourrait s'attendre à ce que de nombreuses combinaisons différentes existent *in-vivo*. En imaginant qu'une portion de chromatine ne peut porter que deux modifications d'histones différentes, avec les 100 marques possibles, il devrait exister 10 000 combinaisons. Pourtant, seule une très faible fraction de toutes ces possibilités est observée *in-vivo*.

Chez *Arabidopsis thaliana*, en analysant la distribution de 12 marques épigénétiques, seules 38 combinaisons différentes ont été observées sur le millier théoriquement possible [Roudier et al., 2011]. De même, chez la drosophile, seules 5 combinaisons sont observées en analysant la distribution de 53 protéines associées à la chromatine [Filion et al., 2010]. Enfin, chez l'homme, seules une quinzaine de combinaisons peuvent être identifiées avec 9 marques d'histones [Ernst et al., 2011].

Ces résultats indiquent que, malgré la très grande diversité de marques d'histones et des complexes responsables de leur mise en place ou de leur lecture, de nombreuses marques se retrouvent associées et établissent ensemble une "identité" à la chromatine. Ceci permet de définir un nombre limité "d'états chromatiniens" correspondant, le plus souvent, à des fonctions différentes.

Malgré la diversité des combinaisons possibles, certaines modifications d'histones sont systématiquement associées, conduisant à un nombre d'états chromatiniens différents assez limité.

1.1.3- De la fibre de 10 nm à la fibre de 30 nm

La répétition de nucléosomes régulièrement espacés conduit à la formation de la fibre de 11 nm ou "beads on a string". L'ADN faisant le lien entre les différents nucléosomes est appelé "ADN de liaison". Cette organisation, qui corres-

1.1 - Organisation structurale et fonctionnelle de la chromatine

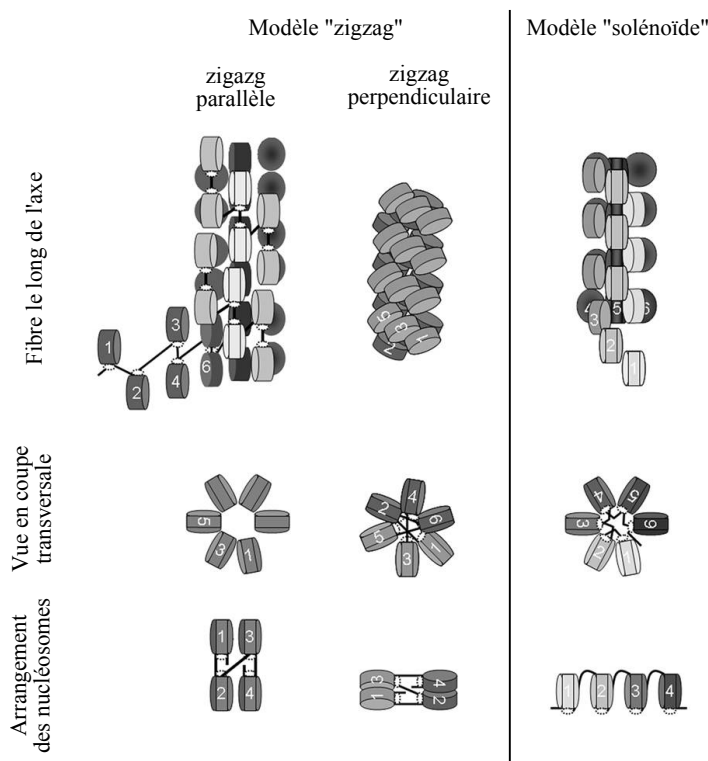


FIGURE 5 – Principaux modèles pour l'organisation de la fibre de 30 nm. Deux familles de modèles sont le plus souvent proposées, celui de type "zigzag" et celui de type "solénoïde" (d'après Szerlong, H. J. and Hansen, J. C. (2011). Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem Cell Biol*, 89(1):24–34 ©2008 Canadian Science Publishing or its licensors. Reproduit avec autorisation).

pond à l'état le plus décompacté de la chromatine, a été observée par microscopie électronique (figure 1B) à faible force ionique (≈ 5 mM).

Lorsque des extractions similaires sont effectuées à plus forte charge ionique, des fibres d'un diamètre de 30 nm sont observées : il s'agit d'un état plus compact de la chromatine, dont la stabilité semble dépendre de la présence de l'histone H1 [Thoma et al., 1979]. L'histone H1 (et son isoforme H5 trouvé dans les érythrocytes des oiseaux) se lie au nucléosome au niveau de l'entrée et de la sortie de l'ADN, influençant donc la conformation de l'ADN de liaison, et donc l'agencement des nucléosomes les uns par rapport aux autres [Syed et al., 2010]. L'assemblage des 8 histones de cœur, de l'ADN et de l'histone de liaison H1 forme un complexe appelé chromatosome.

Plusieurs modèles ont été proposés concernant l'agencement des chromatosomes au niveau de la fibre de 30 nm. Deux grandes familles de modèles existent : l'une basée sur un axe virtuel autour duquel s'enroulent les nucléosomes successifs (modèle "solénoïde"), et l'autre dans lequel l'ADN de liaison traverse l'axe d'enroulement, conduisant à un agencement des nucléosomes en zigzag (modèle "zigzag"). Cette dernière famille peut encore être sous-divisée en plusieurs sous-familles selon la manière dont le zigzag s'effectue (figure 5). Dans le modèle "solénoïde", l'empilement des nucléosomes décrit une simple hélice, alors que dans le modèle "zigzag", ce sont deux hélices d'empilement de nucléosomes qui sont reliées par l'ADN de liaison. Récemment, de nombreuses études ont cherché à déterminer laquelle des deux familles de modèles est la plus proche de la réalité, notamment en jouant sur la taille de l'ADN de liaison. Modifier la taille de l'ADN

de liaison devrait affecter différemment les dimensions de la fibre de 30 nm selon le modèle considéré [Wu et al., 2007]. N'observant pas de relation directe entre ces deux paramètres pour toutes les longueurs d'ADN de liaison testées, il a été proposé que la fibre de 30 nm puisse adopter alternativement la configuration solénoïde ou la configuration zigzag (ces idées ont été discutées dans Szerlong and Hansen [2011]).

La fibre de 30 nm, même si sa structure n'est pas encore bien connue, constitue un niveau additionnel de repliement.

1.1.4- De nombreux indices pour l'existence de boucles géantes de chromatine

La fibre de 30 nm peut également se replier et former des structures dont la topologie est encore moins bien décrite. Ces repliements d'ordre supérieur forment ce que l'on appelle l'organisation à grande échelle de la chromatine, qui repose sur la formation de boucles géantes de chromatine. Les preuves attestant de l'existence de boucles sont nombreuses et diverses.

a- Les chromosomes en écouvillon

Un des premiers indices pour l'existence de boucles de matériel génétique provient sans doute de l'observation des chromosomes en écouvillon. Ces structures apparaissent en fin de méiose dans les ovocytes d'amphibiens et consistent en des boucles de chromatine émanant d'un squelette central. Ces boucles correspondent à des zones de chromatine décondensée et transcriptionnellement active, permettant l'accumulation d'ARNm. La taille de ces boucles a pu être évaluée par microscopie électronique à ≈ 45 kb [León and Kezer, 1990]. Les chromosomes en écouvillon étant visibles par microscopie optique à contraste de phase, ils constituent un indice convaincant pour l'existence de boucles de chromatine.

b- Les chromosomes dépourvus d'histones

Les travaux de Laemmli et collègues en fin des années 1970, portant sur la structure du chromosome en métaphase, ont fourni d'autres observations de boucles géantes de chromatine. En enlevant une grande majorité des protéines chromosomiques (y compris les histones) par compétition avec des polyanions (dextran sulfate et héparine) [Adolph et al., 1977], on obtient un squelette central fibreux entouré d'un halo d'ADN (figure 6A). Ce halo d'ADN présente plusieurs propriétés : (1) il fait en moyenne 10-12 μm de diamètre, (2) il est constitué de boucles d'ADN d'une taille moyenne de 10-30 μm , (3) il est homogène, suggérant que la majorité de l'ADN des chromosomes métaphasiques est organisé en boucles, et (4) la majorité des boucles retournent au squelette protéique en un point très proche de leur point d'origine [Paulson and Laemmli, 1977]. Des images similaires sont observées lorsque les chromosomes sont traités avec 2M NaCl, connu pour enlever les histones.

L'observation ultérieure de coupes transversales de chromosomes métaphasiques [Marsden and Laemmli, 1979] a permis de proposer un modèle dans lequel des

1.1 - Organisation structurale et fonctionnelle de la chromatine

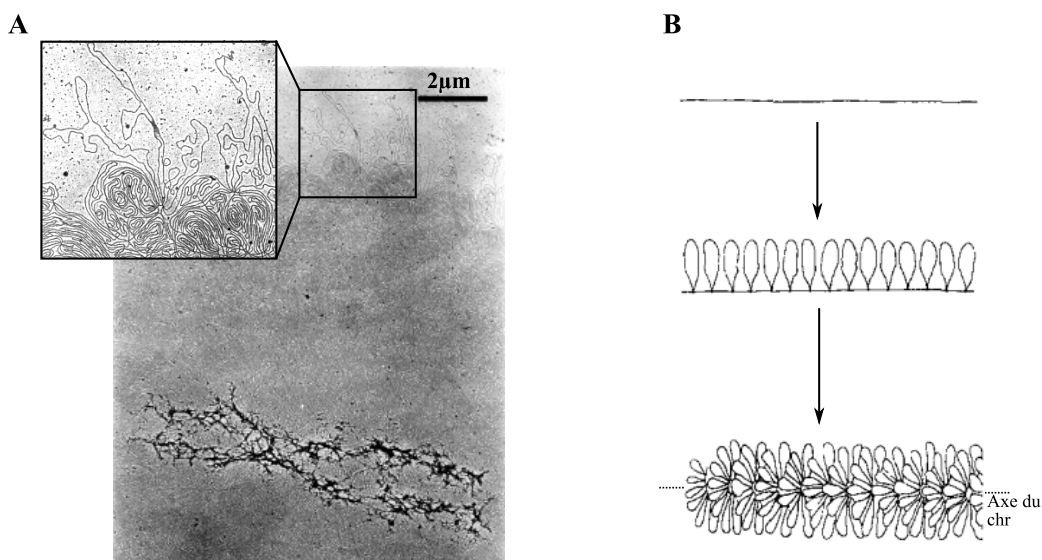


FIGURE 6 – Chromosome en métaphase dépourvu d'histones observé par microscopie électronique. A- Observation des chromosomes, d'après Paulson and Laemmli [1977]. B- Modèle proposé par Marsden and Laemmli [1979] concernant l'organisation de la chromatine du chromosome métaphasique. La fibre se replie pour former des boucles, qui s'organisent ensuite en rosette en s'ancrant sur le squelette du chromosome. Reproduit de Paulson, J. R. and Laemmli, U. K. (1977). The structure of histone-depleted metaphase chromosomes. *Cell*, 12(3):817–828 copyright 1977 et de Marsden, M. P. and Laemmli, U. K. (1979). Metaphase chromosome structure: evidence for a radial loop model. *Cell*, 17(4):849–858 copyright 1979, avec l'autorisation d'Elsevier.

boucles de chromatine, ancrées à leur base, émaneraient radialement d'un squelette protéique formant le cœur du chromosome en métaphase (figure 6B).

c- La matrice nucléaire

Un résidu protéique, appelé "squelette nucléaire", "matrice nucléaire" ou "nucléoïde", est également obtenu lorsque des noyaux sont traités avec des polyanions ou du NaCl [Adolph, 1980]. Autour de ces nucléoïdes, on retrouve également un halo d'ADN. Ces molécules d'ADN du halo sont super-enroulées à cause de l'éviction des nucléosomes [Cook et al., 1976], cette conservation des super-tours constituant une preuve de leur ancrage sur le squelette nucléaire.

Des digestions ménagées par des nucléases ont montré que 3-5% de l'ADN total du noyau est physiquement ancré sur ce squelette nucléaire, le protégeant ainsi de l'action des nucléases [Razin et al., 1979]. L'étude de l'ADN ancré sur le squelette nucléaire a montré qu'il est de nature similaire à celui lié au squelette du chromosome métaphasique (les 2/3 correspondent à des éléments répétés type satellite murin) [Razin et al., 1979]. Cette observation suggère que les points d'ancrage de l'ADN restent constants tout au cours du cycle cellulaire, et donc que la chromatine est également organisée en boucles lors de l'interphase. L'observation de nucléoïdes par microscopie électronique a par ailleurs confirmé l'existence de ces grandes boucles de chromatine [McCready et al., 1979].

Les séquences ancrées sur la matrice nucléaire, qui ont été nommées S/MAR (Scaffold/Matrix Attachment Regions), sont généralement riches en AT [Razin, 2001; Liebich et al., 2002] et correspondent fréquemment à des fragments d'ADN courbes [Fiorini et al., 2006] sujets à des ouvertures locales de la double hélice [Kohwi-Shigematsu and Kohwi, 1990]. Les S/MAR sont retrouvés dans les introns des gènes,

en bordures des régions codantes, ou bien à proximité des séquences régulatrices [Cockerill and Garrard, 1986; Gasser and Laemmli, 1986; Käs and Chasin, 1987; Keaton et al., 2011]. Certains S/MAR possèdent également une fonction d'isolateur en bloquant les effets des enhancers distants [Yusufzai and Felsenfeld, 2004].

Si la nature biochimique de la matrice nucléaire reste sujette à débat, ces observations attestent néanmoins de la présence de boucles de chromatine fermement ancrées.

d- Marche de polymères d'après des hybridations *in-situ*

L'organisation de la chromatine dans le noyau a également été étudiée par hybridation *in-situ* fluorescente (FISH). En rendant deux loci fluorescents, on peut mesurer la distance physique les séparant dans une population nucléaire. En multipliant ces mesures pour différents couples de sondes, on peut ainsi établir une relation reliant la distance génomique (g , en paire de bases) et la distance physique (r , en μm).

Les longues molécules, comme la chromatine, se comporte globalement comme un polymère. Or la physique des polymères prévoit justement une relation entre la distance physique moyenne au carré $\langle r^2 \rangle$, et la longueur de la chaîne. Pour une chaîne gaussienne (marche aléatoire où les monomères n'exercent pas d'interactions entre eux), $\langle r^2 \rangle$ est proportionnel à $N^{2\nu}$, avec $\nu = 1/2$ et N correspondant au nombre de monomères. Dans ce cas de marche aléatoire, on a donc une relation de proportionnalité entre $\langle r^2 \rangle$ et N . Ce modèle simpliste de marche aléatoire peut être complexifié si on ajoute au modèle la notion de "volume exclu" (deux monomères ne peuvent pas être au même endroit dans l'espace, marche aléatoire auto-évitante). Dans ce cas, la relation devient : $\langle r^2 \rangle \propto N^{2\nu}$, avec $\nu \approx 0.588$. D'autres modèles de marche existent, avec chacun leur propre valeur d'exposant ν [Tark-Dame et al., 2011].

Appliqué à la chromatine, la physique des polymères revient à établir la relation reliant $\langle r^2 \rangle$, facilement mesurable par FISH, à la séparation génomique en Mb. Le régime de marche aléatoire a ainsi été observé pour des distances génomiques < 2 Mb [van den Engh et al., 1992], montrant que pour ces distances génomiques, la chromatine se comporte comme une chaîne flexible sans contraintes particulières. Au-delà de 2 Mb, la distance physique tend à devenir indépendante de la séparation génomique, ce qui nécessite d'envisager une marche de polymère contrainte par la présence de boucles.

Les travaux de B.J. Trask ont alors proposé un modèle de boucles géantes de chromatine ancrées périodiquement sur un squelette aléatoire ("Random-walk giant loop" –RWGL–) [Sachs et al., 1995; Yokota et al., 1995]. Ce modèle explique bien le nivellement de la distance physique lorsque l'on augmente la séparation génomique et s'accorde bien avec les observations des chromosomes dépourvus d'histone. Par contre, puisque ce modèle, lorsqu'il est appliqué au chromosome entier, prédit des distances physiques supérieures à la taille du territoire chromosomique, il se révèle inadapté pour comprendre le confinement des chromosomes entiers dans ces territoires discrets (pour la définition de territoire chromosomique, voir page 33). Le RWGL est donc un modèle adéquat pour comprendre le comportement de la chromatine lorsqu'on s'intéresse à des distances largement inférieures à celle du chromosome (figure 7). Dans ce manuscrit (chap 4), nous ferons appel à ce modèle pour des séparations génomiques d'au maximum 15 Mb, soit largement moins que la taille d'un chromosome humain moyen.

1.1 - Organisation structurale et fonctionnelle de la chromatine

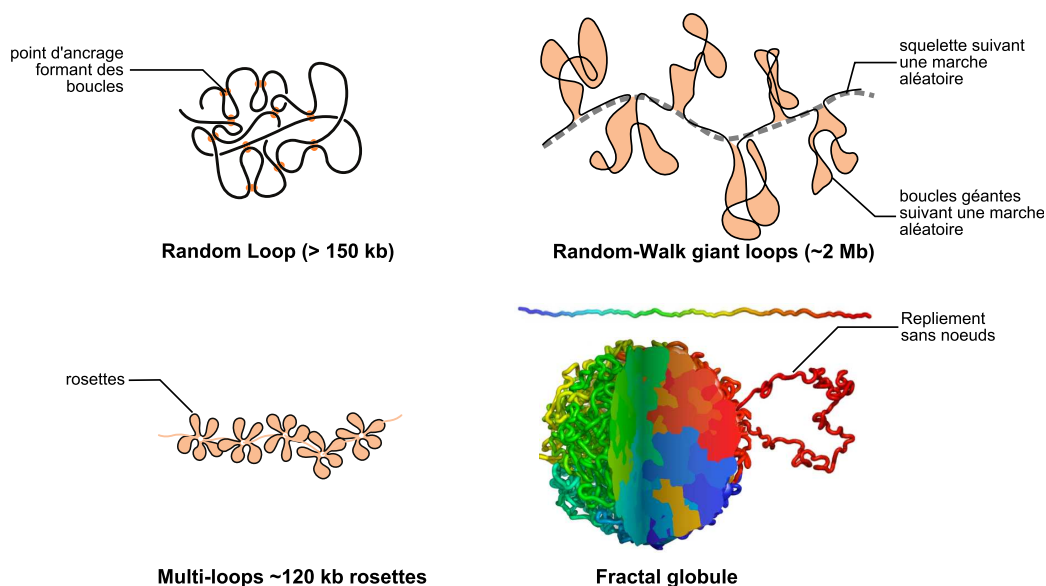


FIGURE 7 – Repliement de la chromatine – modèles "polymères" impliquant des boucles. De haut en bas et de droite en gauche, le modèle "boucles aléatoires" [Bohn et al., 2007], les "boucles géantes ancrées sur un squelette suivant une marche aléatoire" [Sachs et al., 1995], le modèle en "multi-rosettes", [Münkel and Langowski, 1998], et le "globule fractal" [Lieberman-Aiden et al., 2009]. Tous ces modèles incluent la présence de boucles de chromatine pour expliquer l'indépendance entre la distance physique et la séparation génomique pour des distances supérieures à quelques Mb.

D'autres modèles proposent plutôt une organisation de la chromatine en un ensemble de rosettes de 120 kb chacune [Münkel and Langowski, 1998; Münkel et al., 1999]; ou bien en boucles aléatoires de toutes tailles avec des points d'ancrage eux-aussi aléatoires [Bohn et al., 2007]; ou encore une organisation en globule fractal au sein duquel la chromatine ne forme pas de nœuds et est agencée en sous-territoires consécutifs non entremêlés [Lieberman-Aiden et al., 2009] (figure 7). Dans tous ces modèles, qui cadrent relativement bien avec les observations microscopiques au moins à certaines échelles, la marche du polymère a du être contrainte de telle sorte à favoriser la formation de boucles. La présence de ces boucles est absolument nécessaire pour que les prédictions de marches de polymères s'accordent avec les données expérimentales, indiquant qu'elles constituent un niveau de repliement de la chromatine.

L'étude de la matrice nucléaire et la modélisation de marche de polymères d'après les résultats d'hybridation *in-situ* attestent de l'existence de boucles de chromatine. La fibre de 30 nm évoquée dans la sous-section précédente se replie donc en formant des boucles.

1.1.5- Territoires chromosomiques

Ces boucles de chromatine, éventuellement ancrées à un squelette nucléaire, ne sont pas dispersées dans l'ensemble du noyau. En effet, la chromatine de chaque chromosome occupe un territoire nucléaire relativement circonscrit, appelé territoire chromosomique.

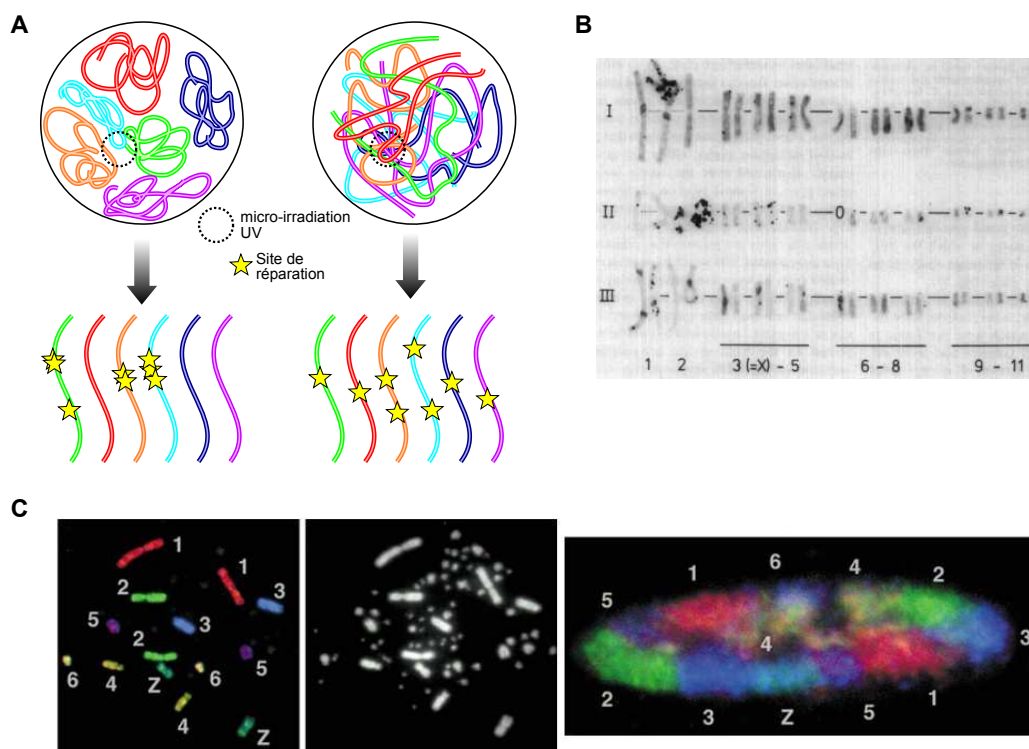


FIGURE 8 – Existence de territoires Chromosomiques. A- Conséquence d'une micro-irradiation laser en fonction de l'organisation des chromosomes dans le noyau. Les sites de réparation sont observés grâce à de la thymidine tritiée. Le nombre de chromosomes ayant subi une ou plusieurs cassures double brin est plus petit si chaque chromosome occupe un territoire nucléaire isolé des chromosomes voisins B- Caryotype obtenu lorsque l'expérience (A) est effectuée sur des fibroblastes de hamster chinois [Zorn et al., 1979]. Ces deux figures sont adaptées de avec autorisation de Macmillan Publishers Ltd : Nature Meaburn, K. J. and Misteli, T. (2007). Cell biology: chromosome territories. *Nature*, 445(7126):379–781. Copyright 2007. C- Marquage fluorescent simultané des macrochromosomes de poulet [Habermann et al., 2001] (à gauche, étalement de métaphase avec le marquage de 7 chromosomes ; au centre, métaphase visualisée avec du DAPI ; à droite, noyau en interphase avec les 7 chromosomes fluorescents). adapté de Habermann, F. A., Cremer, M., Walter, J., Kreth, G., von Hase, J., Bauer, K., Wienberg, J., Cremer, C., Cremer, T., and Solovei, I. (2001). Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Res*, 9(7):569–584, Fig1, copyright 2001, avec l'aimable autorisation de Springer Science and Business Media.

a- Découverte des territoires chromosomiques par micro-irradiation UV

Pour étudier la structure du chromosome en interphase dans le noyau, des cellules de hamster chinois ont été micro-irradiées par un laser UV de manière à créer des cassures dans l'ADN. De la thymidine tritiée est ensuite ajoutée à la culture pendant 2 h, de manière à marquer les portions de chromatine ayant réparé leur ADN. Puis, plusieurs heures après, la distribution de la zone irradiée est analysée sur des noyaux en interphase ou sur des préparations de métaphase.

Le résultat de cette expérience montre (1) que la taille de la zone irradiée change peu au cours du temps, démontrant l'absence de dilution de la chromatine irradiée dans le noyau, et (2) que seuls quelques chromosomes métaphasiques (en moyenne 4) sont radio-marqués [Cremer et al., 1982; Zorn et al., 1979]. Bien qu'indirecte, cette expérience suggère fortement que les différents chromosomes occupent des territoires distincts en interphase, avec assez peu de mélange (figure 8A-B).

b- Chromosome Painting

Les expériences d'hybridation *in-situ* ont permis de confirmer l'existence des territoires chromosomiques (pour un historique des développements technologiques requis pour la préparation des sondes chromosomiques, voir [Cremer and Cremer \[2010\]](#)). Grâce à des sondes couvrant l'intégralité d'un chromosome, il a été possible de visualiser directement le volume occupé par un chromosome en interphase [[Schardin et al., 1985](#)] et de remarquer que celui-ci occupe un territoire discret dans le noyau. Avec l'hybridation *in-situ* fluorescente, il a été possible de visualiser simultanément plusieurs chromosomes d'un eucaryote supérieur [[Cremer et al., 2001](#); [Habermann et al., 2001](#)], et d'analyser leur agencement les uns par rapport aux autres. Il a ainsi été remarqué que chaque chromosome a son propre territoire, chevauchant relativement peu (on discutera ce point plus tard, voir plus loin) celui du voisin [[Habermann et al., 2001](#); [Bolzer et al., 2005](#)]. De plus, les territoires chromosomiques ne sont pas disposés aléatoirement dans le volume nucléaire.

c- Arrangement non-aléatoire des territoires chromosomiques

Il a été observé que les petits chromosomes humains, riches en gènes (chr 16, 17, 19, 20, 21 et 22) ont une position plutôt centrale dans le noyau, alors que les grands chromosomes pauvres en gènes adoptent une position plus périphérique [[Boyle et al., 2001](#); [Tanabe et al., 2002](#)]. Cette tendance générale doit néanmoins être nuancée puisque l'agencement des territoires chromosomiques peut varier d'un type cellulaire à l'autre [[Parada et al., 2004](#)].

Le chromosome humain 18 et 19 ont à peu près la même taille (respectivement 77 et 63 Mb), mais possèdent une densité en gènes complètement différente (3 fois plus de gènes par Mb sur le chr 19 que sur le chr 18). L'observation que le chr 19 occupe une place beaucoup plus centrale que le chr 18 indique que ce serait le critère "densité en gènes" qui contrôlerait l'agencement des chromosomes, plutôt que le critère "taille" [[Croft et al., 1999](#); [Cremer et al., 2003](#)]. Néanmoins, [Bolzer et al. \[2005\]](#) ainsi que [Zeititz et al. \[2009\]](#) n'observent pas de différence significative entre le positionnement du chr 18 et 19 dans des fibroblastes, confirmant que des organisations spécifiques d'un type cellulaire restent possibles.

Cette tendance pour un positionnement des chromosomes en fonction de leur taille/densité en gènes a aussi été observée chez les rongeurs et chez les bovins [[Mayer et al., 2005](#); [Koehler et al., 2009](#)]. Afin de savoir à quel moment dans le développement se mettait en place une telle organisation radiale des chromosomes, [Koehler et al. \[2009\]](#) ont suivi le positionnement du chr 19 bovin (riche en gènes) et du chr 20 bovin (pauvre en gènes) après fécondation *in-vitro*. Ils ont ainsi montré que le positionnement du chr 19 au centre du noyau et du chr 20 en périphérie du noyau ne peut être observé qu'à partir du stade 8-16 cellules, stade à partir duquel débute l'expression du génome zygotique. Ceci indiquerait que le positionnement radial des chromosomes en fonction de leur densité en gène serait relié à la transcription du génome.

d- Imbrication des territoires chromosomiques

Bien que se chevauchant peu, les territoires chromosomiques ne possèdent pas de bordures complètement franches et une certaine imbrication entre les chromosomes est possible [[Branco and Pombo, 2006](#)]. Le volume d'imbrication dé-

pend de la paire de chromosome considérée. Il semble également dépendre de la transcription puisqu'il augmente ou diminue suite à un traitement à l' α -amanitine [Branco and Pombo, 2006]. Combiné à l'agencement non-aléatoire des chromosomes dans le noyau, ces imbrications offrent la possibilité d'établir des contacts entre chromosomes dont on reparlera ultérieurement (page 49).

Ultime niveau de repliement, les boucles de chromatine évoquées dans la sous-section précédente se confinent dans un territoire chromosomique précis. Ces différents territoires sont agencés dans le noyau de telle sorte que les petits chromosomes riches en gènes sont plutôt au centre du noyau, et les grands chromosome plutôt en périphérie. Dans la partie résultat de ce manuscrit, nous ferons appel à cette observation (paragraphe 3.3.2 et paragraphe 5.4).

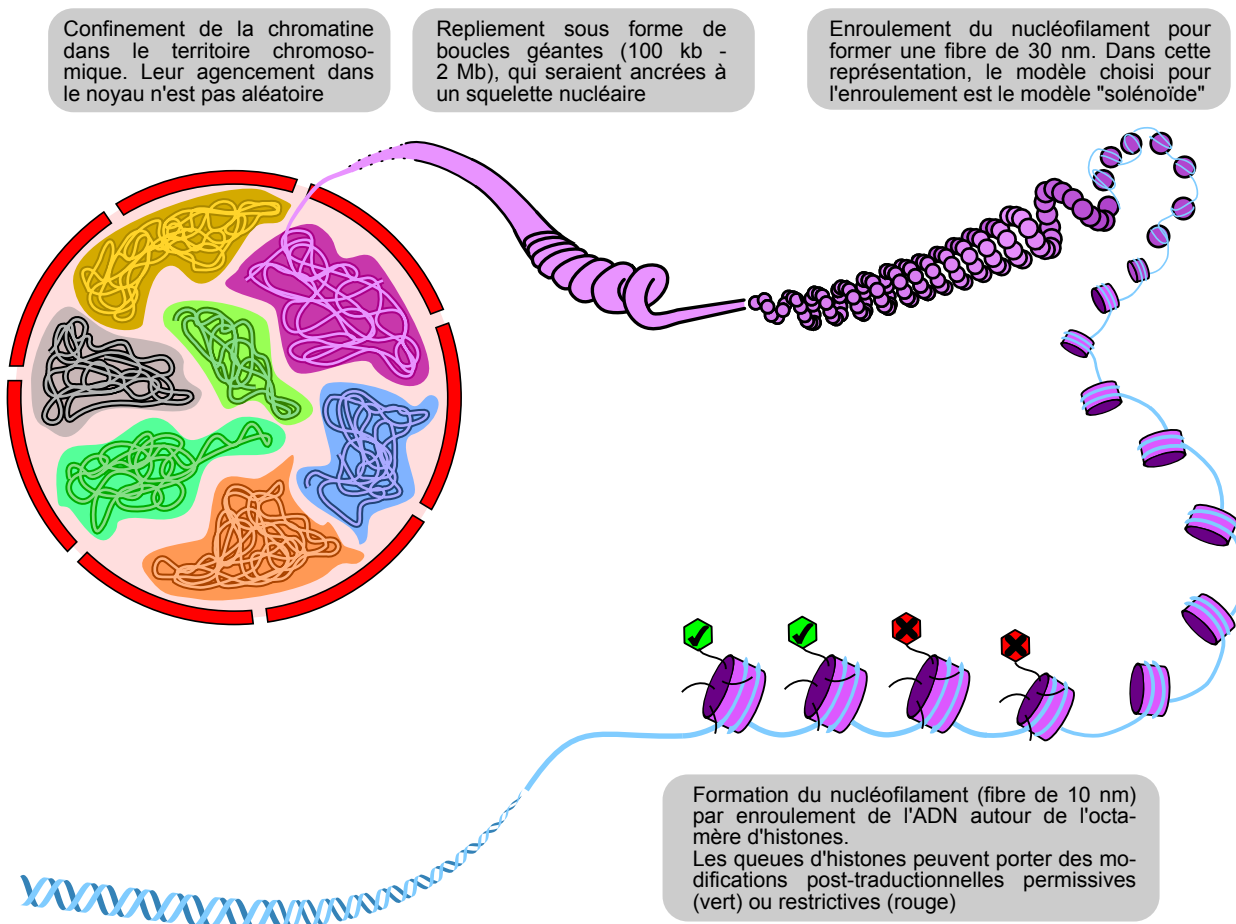


FIGURE 9 – Repliement hiérarchique de la chromatine. La première étape de repliement consiste en l'enroulement de l'ADN autour de l'octamère d'histone pour former le nucléofilament, lequel s'enroule en une fibre de 30 nm (ici représentée sous forme de solénoïde). Celle-ci se replie à nouveau pour former des boucles de chromatine vraisemblablement ancrées à la matrice nucléaire. Enfin, l'ensemble est confiné dans un volume défini du noyau –le territoire chromosomique–.

Conclusion

L'organisation de la chromatine est donc hiérarchique. A l'échelle de la centaine de paires de base, l'ADN s'enroule autour de l'octamère d'histone, formant des nucléosomes et la fibre de 11 nm. Cette fibre peut se compacter en une fibre de 30 nm dont la structure n'est pas encore complètement comprise. A l'échelle de la dizaine de kilobase à quelques mégabases, la chromatine forme des boucles qui seraient, selon les modèles classiques, ancrées à leur base. Enfin, à l'échelle du chromosome entier (jusqu'à 250 Mb chez l'homme), la chromatine se confine dans un territoire discret de quelques μm , isolé (en tout cas en première approche) des autres chromosomes. Cette organisation hiérarchique est résumée dans la figure 9.

Ces différents niveaux de repliement vont avoir un effet sur les fonctions nucléaires. Dans la section suivante, nous verrons que ces repliements autorisent la formation de contacts spécifiques entre différentes portions de chromatine, et que ces contacts sont nécessaires à l'expression de certains gènes.

Si la réplication de l'ADN est pour l'instant mise de côté, nous verrons dans le chapitre 2 que des contacts préférentiels peuvent également s'établir entre régions répliquées de manière synchrone.

1.2- Des repliements contrôlés et dynamiques

Dans cette partie, j'aborderai les différentes expériences et résultats ayant permis d'appréhender la dynamique de la chromatine et l'adéquation entre sa structure (repliement) et la fonction sous-jacente. L'accent sera principalement placé sur la structure de la chromatine à grande échelle (de quelques dizaines de kilobases à plusieurs méga-bases, voire le chromosome entier). Mais, pour démarrer cette section, je détaillerai les méthodes de la famille 3C qui permettent d'analyser comment la chromatine se replie dans le noyau. Puisque cette méthodologie sera très utilisée dans la partie résultat de ce manuscrit, je discuterai de leurs limites et champs d'application.

1.2.1- Méthodes de la famille 3C

a- Aperçu des différentes méthodes de la famille 3C

Les méthodes de la famille 3C (Chromosome Conformation Capture) [Dekker et al., 2002] ont récemment émergé afin d'étudier comment la chromatine se replie. Elles reposent toutes sur la capture de contacts chromatinien préférentiels dans une population cellulaire après à une fixation (cross-link) au formaldéhyde. Puis, la chromatine cross-linkée est digérée à l'aide d'une enzyme de restriction, une ligation est effectuée de manière à rabouter ensemble les portions de chromatine reliées par le cross-link, puis l'ADN est purifié (figure 10). A la fin de cette étape, on se retrouve donc avec des molécules d'ADN chimériques qui reflètent les relations de proximité dans l'espace nucléaire. La manière d'analyser ces molécules d'ADN chimériques va différer d'une sous-méthode à l'autre.

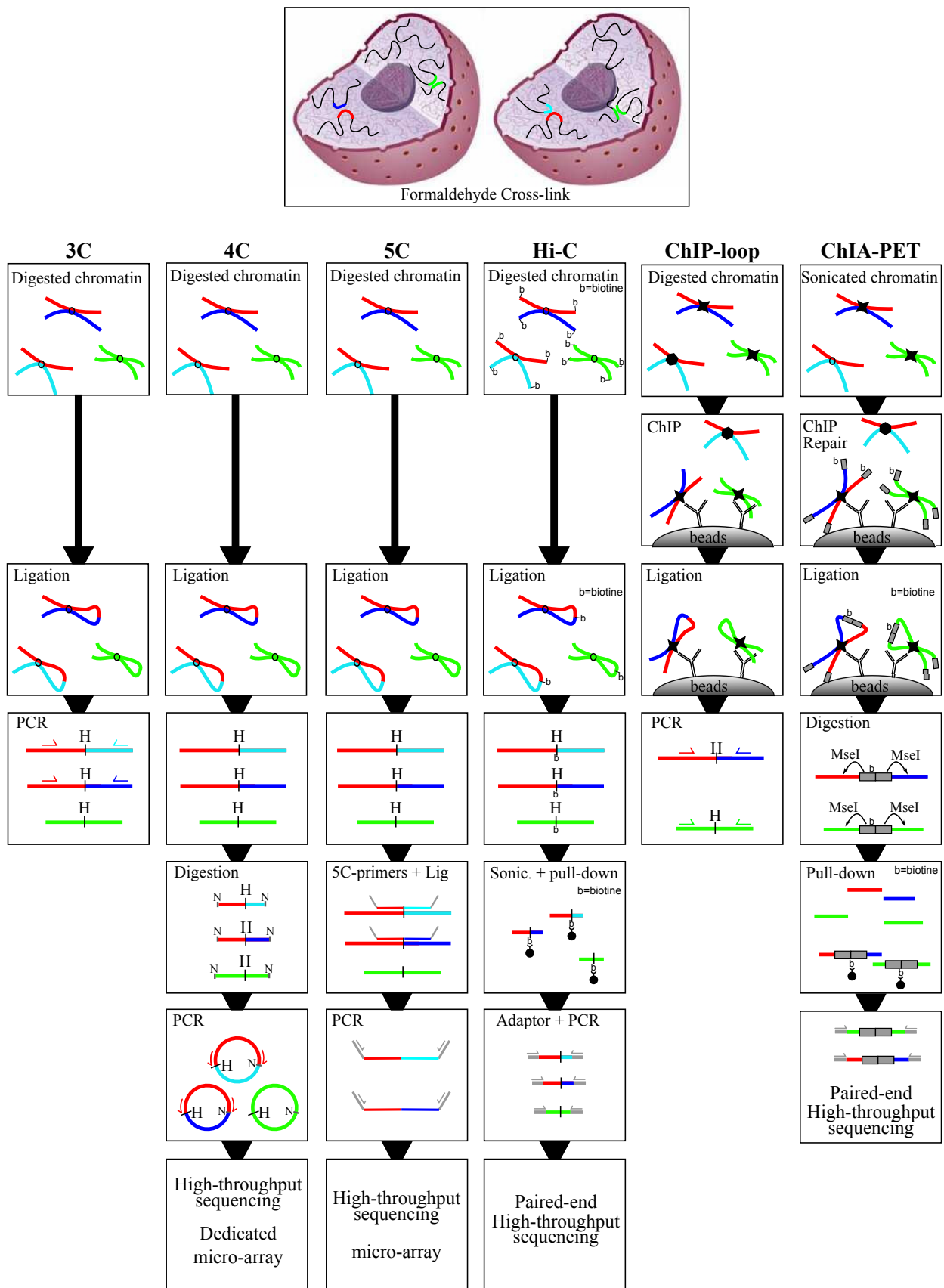


FIGURE 10 – Aperçu des différentes méthodes de la famille 3C. Voir le texte pour les détails.

1.2- Des repliements contrôlés et dynamiques

Dans la méthodologie 3C, les fragments d'ADN chimériques sont analysés par PCR (figure 10). L'abondance du produit PCR reflètera l'abondance du produit de ligation, et donc la proximité des deux loci dans le noyau. L'abondance du produit de ligation peut être évaluée par PCR semi-quantitative (lecture de densité optique sur gel d'agarose) [Dekker et al., 2002] ou par PCR quantitative [Hagège et al., 2007]. Dans les deux cas, l'efficacité de la PCR doit être normalisée pour standardiser les différences d'amplification d'un couple d'amorces à l'autre [Dekker et al., 2002; Dekker, 2006]. D'autre part, la probabilité qu'une interaction entre deux loci soit capturée est directement reliée à la distance génomique les séparant : plus ils sont proches (en kb), plus le cross-link entre eux aura de chance d'être fréquent. Ainsi, si A-B-C-D-E-F sont 6 fragments de restrictions, une ligation A-D a plus de chance de survenir qu'une ligation A-F si le repliement de la chromatine est aléatoire. En conséquence, pour détecter une interaction préférentielle, il ne suffit pas de se baser sur la capacité à effectuer une ligation particulière. Pour prouver qu'une interaction préférentielle existe entre A et D, il sera par contre convaincant de montrer que la ligation A-D survient *plus souvent* que la ligation A-C. Une telle observation sera la preuve que le repliement n'est pas uniquement dirigé par des collisions aléatoires [Dekker et al., 2002; Dekker, 2006].

Dans la méthode 3C, un *a priori* sur l'existence d'interactions est nécessaire, puisque leur détection nécessite un couple d'amorces sur chaque partenaire. La méthode 4C (Circular Chromosome Conformation Capture) permet de s'affranchir de cet *a priori* en analysant tous les partenaires interagissant avec un locus donné (appelé "point de vue") [Simonis et al., 2006; Zhao et al., 2006]. Pour cela, des étapes expérimentales additionnelles sont nécessaires : dans le protocole développé par Simonis et al. [2006], une digestion supplémentaire est effectuée à l'aide d'une enzyme reconnaissant 4 nucléotides, suivi d'une circularisation. Enfin, une PCR inverse à l'aide d'amorces au niveau du point de vue permet d'amplifier toute séquence ayant été capturée dans les étapes préalables. Ainsi, tous les partenaires interagissant avec le point de vue sont amplifiés par une unique PCR, et peuvent être analysés par puce à ADN ou séquençage haut-débit (figure 10). En somme, cette méthode permet de mesurer les interactions d'un locus avec tout le génome.

D'autres méthodes à haut-débit ont également été développées. Le Hi-C analyse directement les molécules chimériques résultant de la première ligation. Ainsi, en ne sélectionnant que les zones de recombinaison, et en les analysant par séquençage haut-débit (figure 10), la méthode Hi-C permet d'analyser l'ensemble des interactions du noyau, tout le génome avec tout le génome [Lieberman-Aiden et al., 2009].

Le 5C (3C-Carbon Copy) adopte une stratégie différente. L'ADN chimérique est hybridé avec un mélange d'oligonucléotides ciblant exactement les sites de restrictions [Dostie et al., 2006]. Les oligonucléotides utilisés sont dessinés de telle sorte que, lorsque leurs deux extrémités 3' se retrouvent en face, ils puissent subir une ligation par la Taq DNA ligase. Ainsi, une copie-carbone des événements de ligation est créée grâce aux oligonucléotides. L'extrémité 5' de ces oligonucléotides contient également la séquence des amorces T3 et T7, permettant, en une seule PCR, d'amplifier la copie-carbone. Cette PCR est enfin analysée par séquençage haut-débit ou puce à ADN (figure 10).

Des méthodes analogues ont également été développées pour coupler la méthodologie 3C avec celle de la ChIP, permettant ainsi de détecter des interactions basées sur un facteur commun. Le ChIP-loop [Horike et al., 2005] et le ChIA-PET [Fullwood et al., 2009] sont respectivement les versions du 3C et du Hi-C incluant une étape préalable de ChIP (figure 10).

Pour plus de détail sur cette méthodologies de la famille 3C, je vous invite à consulter la revue écrite par [de Wit and de Laat \[2012\]](#).

b- Discussion des champs d'application et limites du 4C et du Hi-C

Comme les méthodologies 4C et Hi-C ont été largement utilisées dans la partie résultat de ce manuscrit, une discussion sur leurs limites et leurs champs d'application peut d'ores et déjà être engagée.

4C La résolution des expériences 4C dépend en premier lieu de l'enzyme de restriction utilisée et de sa fréquence de coupure. HindIII est l'enzyme la plus utilisée [[Simonis et al., 2006](#); [Würtele and Chartrand, 2006](#); [Palstra et al., 2008](#); [Simonis et al., 2009](#); [Splinter et al., 2011](#); [Noordermeer et al., 2011a](#); [Hakim et al., 2011](#)], même si d'autres enzymes de restrictions à 6 nucléotides ont également été employées (EcoRI [[Tolhuis et al., 2011](#)] et BglII [[Schoenfelder et al., 2010b](#)]). Ces enzymes coupant classiquement tous les 4 kb, il est donc impossible d'analyser les interactions avec une plus grande résolution.

Ceci est d'autant plus vrai que les données 4C brutes sont, le plus souvent, lissées à l'aide d'une moyenne glissante ou d'une médiane glissante. Ce post-traitement permet d'atténuer les différences d'amplification par PCR d'un fragment de restriction à son voisin (pour une comparaison des données brutes et des données lissées, voir [Simonis et al. \[2006\]](#) et [de Wit and de Laat \[2012\]](#)). Ce lissage, classiquement effectué sur quelques dizaines de kb [[Simonis et al., 2006](#); [Palstra et al., 2008](#); [Noordermeer et al., 2011a](#); [Tolhuis et al., 2011](#)], empêche la détection d'interactions très fines, typiquement entre un promoteur et un enhancer. Pour de telles interactions, l'utilisation du protocole 3C sera vraisemblablement plus adaptée. La méthodologie 4C est au final nettement plus adaptée à la détection d'interactions à longue-distance (> 100 kb).

Des stratégies employant des enzymes de restriction coupant plus fréquemment ont pu être développées (MseI [[Zhao et al., 2006](#)], NlaIII [[Noordermeer et al., 2011b](#)] ou DpnII [[Lower et al., 2009](#); [Bantignies et al., 2011](#); [Montavon et al., 2011](#)]). L'utilisation de ces enzymes, coupant en moyenne tous les 250 pb, devrait en théorie permettre d'augmenter la résolution de la méthodologie 4C. Cela dit, en appliquant ce protocole à des cellules humaines [[Lower et al., 2009](#)] ou murines [[Noordermeer et al., 2011b](#)], la PCR 4C effectuée sur 200 ng d'ADN (soit ≈ 60000 génomes haploïdes) ne permet d'amplifier au maximum que ≈ 60000 produits de ligation. Ce nombre restreint de produits de ligation est à comparer avec les 10^7 de sites NlaIII ou DpnII du génome. Essayer d'augmenter la résolution en utilisant des enzymes à coupure fréquente conduit donc à un faible pourcentage de sites NlaIII/DpnII capturés, et se fait au détriment de l'analyse des interactions plus longues distances.

Hi-C Le Hi-C, en fournissant des matrices d'interactions, présente l'énorme avantage de permettre l'analyse de toutes les interactions deux-à-deux [[Lieberman-Aiden et al., 2009](#)]. Ainsi, en une seule expérience, l'ensemble de la conformation du noyau humain peut en théorie être analysé. Cependant, comme pour le 4C utilisant des enzymes à coupure fréquente, la principale limite du Hi-C réside dans ce ratio entre détection réelle d'événements de ligation et le nombre théorique de recombinaisons différentes possibles. La profondeur de lecture du Hi-C

1.2- Des repliements contrôlés et dynamiques

(quelques dizaines de millions de lectures [Lieberman-Aiden et al., 2009]) ne permet pas de couvrir la combinatoire (10^6 sites de restrictions dans le génome humain $\Rightarrow 10^{12}$ recombinaisons possibles). De ce fait, cette méthode apporte une vision globale des partenaires, ne permettant que rarement la visualisation directe des profils d'interaction d'un locus donné le long du chromosome (cette limite du Hi-C sera discutée en fin de chapitre 2, et illustrée figure 24).

L'application de cette méthodologie chez des organismes à plus petits génomes (avec un protocole légèrement modifié [Duan et al., 2010; Tanizawa et al., 2010] et éventuellement combinée à une analyse plus méticuleuse [Yaffe and Tanay, 2011; Sexton et al., 2012]) a permis d'améliorer la résolution. Ainsi, chez la levure (génomme ≈ 12 Mb), une résolution de l'ordre du kb a pu être atteinte [Duan et al., 2010; Tanizawa et al., 2010] et chez la drosophile (génomme ≈ 120 Mb), une résolution de 100 kb a pu être atteinte pour les interactions entre chromosomes [Sexton et al., 2012].

Comme le mentionne Sexton et al. [2012], ces questions de résolution dépendent énormément de la distance génomique à laquelle on se place. Pour des loci séparés de quelques centaines de kb, la résolution sera bien meilleure (dans leur cas, ≈ 10 kb) que pour des loci situés de part et d'autre du même chromosome, voire situés sur des chromosomes différents (dans leur cas, ≈ 100 kb). Ceci signifie que les données Hi-C humaines [Lieberman-Aiden et al., 2009] sont une mine d'information pour le comportement globale des interactions de quelques mégabases. Dans ce manuscrit, ces données Hi-C seront utilisées à ce propos (figure 46 dans le chapitre 4).

Une vision des contacts dans une population cellulaire Quelle que soit la méthode de la famille 3C utilisée, les contacts capturés reflètent une vision statistique des interactions survenant dans une population cellulaire. Le fait de capturer une interaction $A \leftrightarrow B$ ne signifie pas que A et B interagissent dans toutes les cellules. De même, capturer deux interactions $A \leftrightarrow B$ et $A \leftrightarrow C$ n'implique pas forcément que B et C interagissent.

De fait, pour modéliser la structure du noyau qui expliquerait au mieux les matrices de contact de type Hi-C, certaines stratégies ne cherchent pas à établir une configuration moyenne typique, mais plutôt 10000 configurations différentes incorporant chacune un certain nombre de contacts chromatiniens et reflétant chacune la diversité cellulaire [Kalthor et al., 2012]. Les auteurs arrivent ainsi, dans leurs 10000 configurations de noyaux, à reproduire le positionnement des territoires chromosomiques observé par FISH.

Cette famille de méthode permet de capturer, dans une population cellulaire, les contacts entre portions de chromatine. Chaque méthode a son champ d'application : la 3C à courte de distance, la 4C et le Hi-C à plus longue distance. Pour l'heure, ces méthodes ont été essentiellement utilisées pour capturer des interactions en lien avec la transcription (contact enhancer - gènes, contact entre éléments régulateurs) que nous allons détailler dans la suite de ce chapitre.

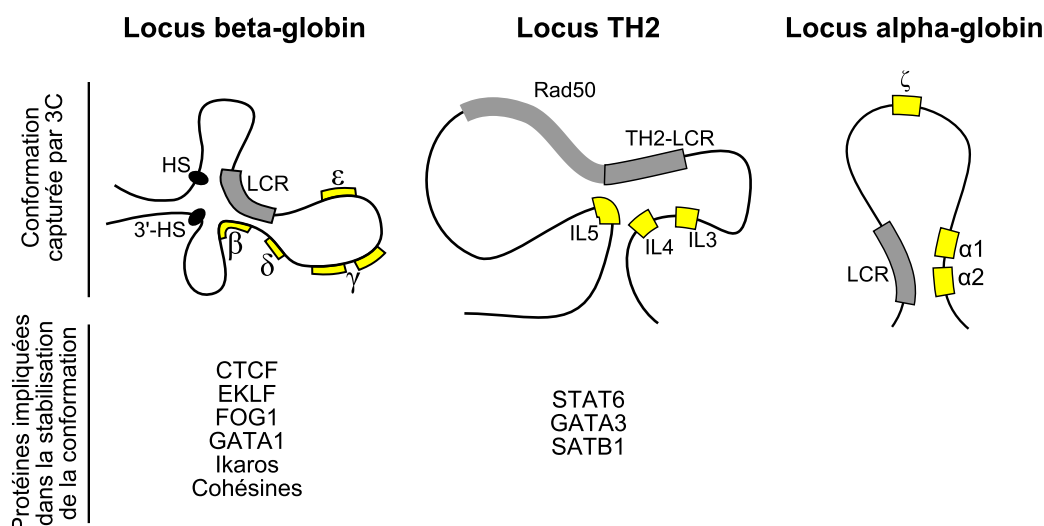


FIGURE 11 – Contacts capturés par 3C entre des enhancers et leurs gènes cibles. A gauche : conformation du locus β -globine dans les cellules exprimant ce locus. Dans ces cellules, les sites hypersensibles à la DNase et le LCR forment ensemble un micro-environnement ("chromatin hub") au sein duquel est recruté le gène exprimé du locus (adapté de [Patrinis et al. \[2004\]](#)). Au centre, conformation du locus TH2 chez les lymphocytes T. L'extrémité 3' de Rad50, qui correspond au LCR du locus, interagit avec IL3, IL4 et IL5 (adapté de [Spilianakis and Flavell \[2004\]](#)). A droite, conformation du locus α -globine chez les précurseurs érythroïdes humains. Le LCR interagit avec les deux gènes $\alpha 1$ et $\alpha 2$.

1.2.2- Des boucles de chromatine locales en lien avec leurs fonctions

Les méthodologies de la famille 3C ont permis de détecter de nombreux types de contacts chromatinien, la plupart en lien avec l'activation ou la répression génique. Dans cette partie, nous détaillons les types de contacts que ces méthodologies peuvent capturer.

a- Contacts enhancer – gène

Les éléments régulateurs de la transcription peuvent parfois être localisés à plusieurs dizaines de kb de leurs gènes cibles. Pour expliquer ces effets à longue distance, différents modèles ont été proposés, dont certains impliquent la présence des boucles de chromatine rapprochant dans l'espace séquences régulatrices et gènes cibles, tandis que d'autres impliquent un transfert de co-activateurs le long de l'ADN [[Bulger and Groudine, 1999](#)]. L'observation de contacts préférentiels au niveau du locus β -globine entre le LCR (Locus Control Region, figure 11) et le gène globine exprimé a largement conforté les modèles impliquant la formation de boucles de chromatine. Ces contacts préférentiels ont été détectés par deux méthodes différentes : le RNA-TRAP [[Carter et al., 2002](#)] et la 3C [[Tolhuis et al., 2002](#)]. Ce contact est extrêmement spécifique et contrôlé puisque les cellules n'exprimant pas de β -globine ne présentent pas de contact [[Palstra et al., 2003](#)]. De plus, les contacts enhancer – gènes changent de l'embryon au fœtus en fonction de l'activation séquentielle des gènes globines [[Palstra et al., 2003](#)].

De la même façon, un contact entre le LCR du locus α -globine et les gènes $\alpha 1$ et $\alpha 2$ a également été observé chez des précurseurs érythroïdes humains, ainsi que chez des souris portant le locus α -globine humain [[Vernimmen et al., 2009](#)] (figure 11).

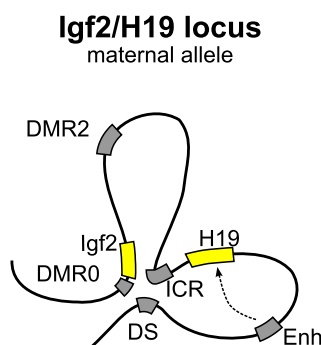


FIGURE 12 – Contact entre les régions différemment méthylées du locus H19/Igf2. Sur l'allèle maternel, la région différemment méthylée DMR0, la région de contrôle de l'empreinte (ICR) et le site de fixation de CTCF situé en aval (DS) s'agrègent. Les boucles formées permettent alors à l'enhancer (Enh) d'activer l'expression de H19. DMR0, DMR2, ICR et DS ne sont pas méthylés sur l'allèle maternel.

D'autres contacts enhancer – gène ont également été observés au niveau du locus TH2, un site d'expression des cytokines des lymphocytes T. Chez ces lymphocytes T, le LCR du locus TH2 (qui est contenu dans le gène de réparation de l'ADN Rad50) peut interagir avec les interleukines IL3, IL4 et IL13, formant une double boucle (figure 11). Cette interaction entre le LCR et les cytokines est spécifique du lignage T et semble se renforcer lors de la différenciation en lymphocyte TH2 qui s'accompagne de l'expression de ces trois cytokines [Spilianakis and Flavell, 2004]. Les facteurs SATB1, STAT6 et GATA3 semblent à la fois essentiels pour l'expression des cytokines et pour l'établissement de cette conformation tridimensionnelle particulière [Spilianakis and Flavell, 2004; Cai et al., 2006].

b- Contacts silenceur – gène

Les complexes polycomb (PcG) sont connus pour réprimer l'expression de gènes Hox en se fixant sur des éléments de réponse au Polycomb (PRE). Par 3C, il a été démontré que des interactions entre PRE et gènes cibles existent chez la drosophile, favorisant probablement ainsi la répression génique [Lanzuolo et al., 2007]. La répression dirigée par les complexes Polycomb, y compris chez les mammifères, ainsi que les aspects tridimensionnel associés ont été récemment discutés [Bantignies and Cavalli, 2011].

c- Contacts entre régions régulatrices

Chez les précurseurs érythroïdes, la plupart des sites hypersensibles à la DNase du locus β -globine peuvent s'agréger, formant ensemble un micro-environnement dans le noyau appelé "Active Chromatin Hub". Les gènes du locus globine sont ensuite recrutés au sein de ce micro-environnement, de manière synchrone avec leur activation au cours du développement [Palstra et al., 2003; Patrinos et al., 2004; de Laat and Grosveld, 2003]. De nombreux facteurs sont impliqués dans la formation et le maintien de ce micro-environnement (EKLF [Drissen et al., 2004], GATA1 et FOG1 [Vakoc et al., 2005], ainsi qu'Ikaros [Keys et al., 2008]).

De la même façon, au niveau du locus Igf2-H19, qui s'étend lui aussi sur approximativement 100 kb, un regroupement des éléments régulateurs du locus a été observé. En effet, les régions différemment méthylées contrôlent la conformation du locus de telle sorte que l'enhancer soit éloigné d'Igf2 sur l'allèle maternel

[Murrell et al., 2004; Qiu et al., 2008]. Il a par ailleurs été montré que CTCF [Kurukuti et al., 2006] et les complexes cohésines [Nativio et al., 2009] sont nécessaires pour mettre en place cette conformation tridimensionnelle.

Les éléments de réponses à Polycomb (PRE) aussi s'associent dans le noyau [Lanzuolo et al., 2007; Bantignies et al., 2011]. Ils contribuent à créer un micro-environnement répressif, qui se manifeste par les corps nucléaires Polycomb dans le noyau ("Polycomb body"), au niveau desquels sont recrutés les gènes réprimés.

Des contacts locaux peuvent être détectés par les méthodes 3C entre régions régulatrices, ou entre région régulatrice et gènes cibles. Ces boucles de chromatine ont des rôles majeurs, aussi bien au niveau de l'activation que de la répression génique. La distance typique des contacts évoqués jusqu'alors est de l'ordre de la centaine de kb. Néanmoins, des contacts à plus longues distances existent et sont détectés par des approches 4C/5C/Hi-C.

1.2.3- Repliement et définition de domaines chromatinien

Un gène en position ectopique peut se retrouver exprimé dans une fraction des cellules (variégation par effet de position) ou bien voir son niveau d'expression amoindrie (effet de position stable) [Yokoyama et al., 1990; Dobie et al., 1996]. Cette observation atteste de l'importance du contexte chromatinien pour comprendre la régulation génique. Cette question de contexte (ou domaine, ou environnement) chromatinien sera discutée dans cette sous-section.

a- Segmentation de domaines fonctionnels d'après les profils d'expression et de modifications épigénétiques

L'analyse des données transcriptomiques montre que le génome peut être sub-divisé en des régions soit riches, soit pauvres en gènes exprimés. Cette tendance a été observée chez la drosophile, la souris et l'homme [Caron et al., 2001; Spellman and Rubin, 2002; Versteeg et al., 2003; Mijalski et al., 2005]. Chez l'homme, les régions riches en gènes exprimés ont été appelées "Ridges" (Regions of Increased Gene Expression) et les régions pauvres en gènes exprimés "Anti-Ridges" [Versteeg et al., 2003]. Ces Ridges, qui sont également riches en GC, denses en SINE et pauvres en LINE, constituent une première segmentation du génome en domaines fonctionnels (figure 13A).

Un autre niveau de segmentation a été observé avec les profils de modification d'histones. Une large portion du génome peut ainsi être divisée en fragments denses ou pauvres en H3K9me2 [Wen et al., 2009]. De même, les profils en H3K27me3 [Pauler et al., 2009] et H3K5ac [Cuddapah et al., 2009] segmentent le génome en une succession de régions enrichies ou appauvries en ces marques. Ces domaines fonctionnels identifiés par les modifications d'histones possèdent plusieurs propriétés : (1) ils sont le plus souvent spécifiques d'une lignée cellulaire ; (2) ils sont conservés entre les régions synténiques de la souris et l'homme [Wen et al., 2009] ; et (3) des sites de fixation de CTCF sont observés à la jonction d'un domaine à l'autre [Cuddapah et al., 2009]. Ces segments aux niveaux de modifications épigénétiques similaires laissent donc envisager un niveau de régulation

1.2- Des repliements contrôlés et dynamiques

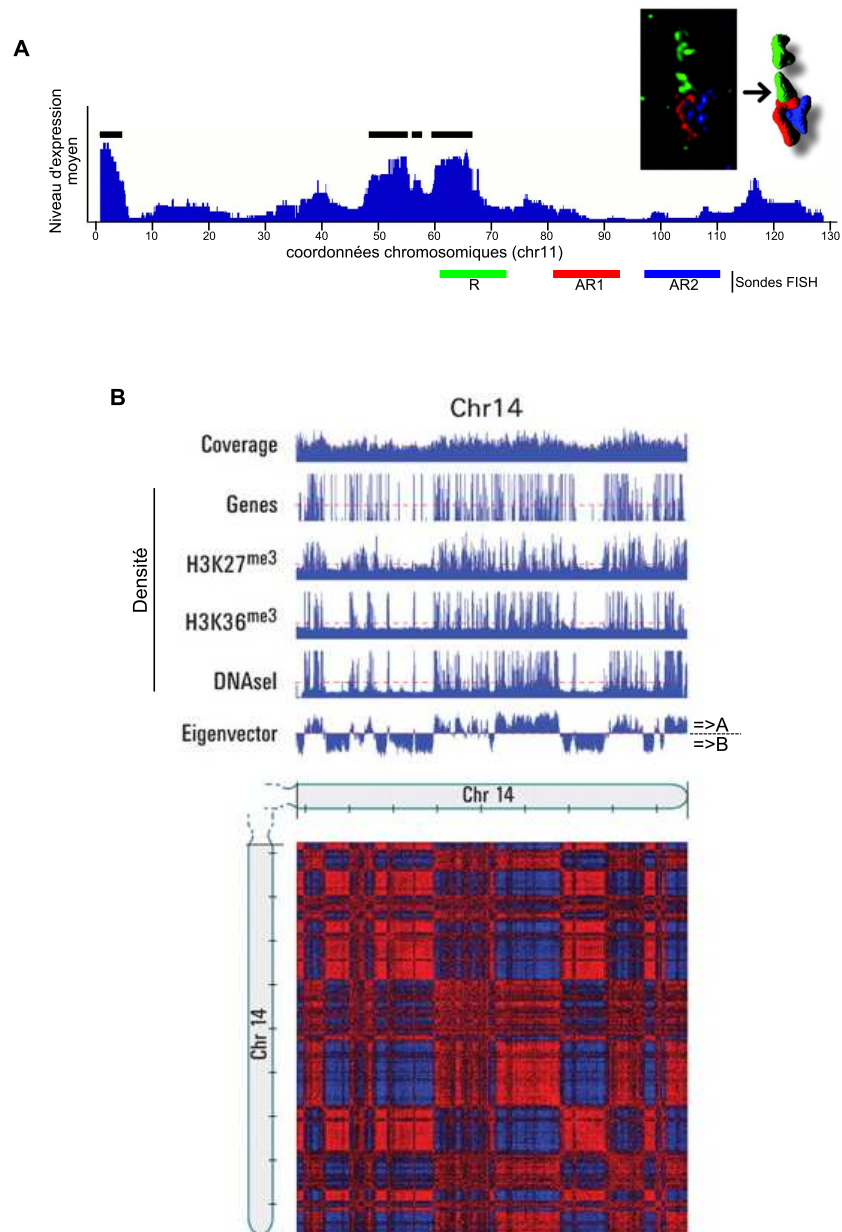


FIGURE 13 – Existence de domaines chromatinien.

A- Niveau d'expression moyen dans plusieurs tissus [Caron et al., 2001]. Le profil montré, qui correspond au chromosome 11 humain, a été obtenu à l'aide d'une médiane glissante sur les profils d'expression de 8 tissus. Les régions riches en gènes exprimés ("Ridges") sont indiquées par des traits noirs. L'étude ayant été effectuée sur la cartographie génétique de 1999, les coordonnées le long du chromosome 11 sont approximatives. Reproduit avec autorisation de AAAS. Trois sondes FISH (mélange de plusieurs BAC) ont été utilisées pour analyser l'agencement de "Ridge" (R) et des "Anti-Ridge" (AR). Une des images obtenues [Goetze et al., 2007] (en haut à droite) montre l'absence de recouvrement entre "Ridge" et "Anti-Ridge". Reproduit avec autorisation.

B- Matrice de corrélation Hi-C. Une corrélation de Pearson est effectuée à partir des matrices d'interactions normalisées pour identifier les loci interagissant avec les mêmes partenaires [Lieberman-Aiden et al., 2009]. Cette analyse a révélé l'existence de deux compartiments (A et B, eigenvector), codés avec une table de correspondance bleu-rouge. Au-dessus de la matrice de corrélation sont représentées les densités en gènes, en H3K27me3, en H3K36me3 et en sites hypersensibles à la DNase. La couverture en lectures est également indiquée (Coverage). Reproduit avec autorisation de AAAS.

de la chromatine qui dépasse largement le cadre d'un seul gène, et intègre sur quelques centaines de kb la contribution de plusieurs régions transcrites.

Cette organisation de la chromatine en segments (domaines) a une influence sur le niveau d'expression. En analysant 90 clones présentant chacun une intégration unique d'un transgène GFP, [Gierman et al. \[2007\]](#) ont en effet observé que le niveau d'expression du transgène est plus fort lorsque celui-ci est intégré au sein d'un "Ridge". Plus surprenant, ils ont noté que le niveau d'expression du transgène corrèle moins avec celui des deux gènes adjacents, qu'avec celui du niveau global d'expression du "Ridge". En somme, un transgène fortement exprimé et inséré dans un "Ridge" peut tout à fait avoir comme voisin un gène réprimé. Ces observations démontrent clairement que le contrôle de l'expression génique ne dépend pas uniquement de la conformation locale de la chromatine : le domaine chromatinien entier joue indubitablement un rôle clé.

b- Segmentation en domaines structuraux

En plus de la segmentation de la chromatine en domaines fonctionnels, une segmentation structurale a également été observée. Par FISH, les "Ridge" et "Anti-Ridge" occupent des territoires distincts dans le noyau, avec très peu de chevauchement [[Goetze et al., 2007](#)] (figure 13A). La chromatine serait donc organisée de telle sorte que les domaines fonctionnels, identifiés à partir du niveau de transcription ou du niveau de modifications épigénétiques, correspondent à des sous-compartiments distincts dans le noyau.

Cette délimitation en domaines structuraux est aussi observée par les méthodes de la famille 3C. En effet, par 4C, il a été montré que le locus *HoxD* est agencé en deux sous-compartiments relativement isolés au niveau du tronc antérieur de souris. L'un de ces sous-compartiments contient les gènes inactifs (côté centromère) tandis que l'autre comprend les gènes actifs (côté télomère). Ainsi, en prenant le gène actif *Hoxd4* comme point de vue, [Noordermeer et al. \[2011b\]](#) ont montré qu'il interagit avec les autres gènes actifs voisins (*Hoxd3* et *Hoxd1*), mais très peu avec les gènes inactifs. A l'inverse, un point de vue 4C situé au sein d'un gène inactif (*Hoxd13*) interagit avec les autres gènes inactifs (*Hoxd11*, *Hoxd9*), mais très peu avec les gènes actifs.

Une correspondance entre domaine fonctionnel et domaine structural a également été observée chez la drosophile au niveau du locus *Bithorax* (BX-C). Ce locus d'environ 500 kb est réprimé par les complexes Polycomb et est entouré d'euchromatine. Une expérience 4C, effectuée lorsque la répression par les complexes Polycomb est déjà effective, a montré qu'un site de liaison des complexes Polycomb, situé au centre du locus BX-C, s'associe fréquemment avec les 3 gènes réprimés du locus, mais très peu avec les gènes exprimés situés immédiatement en 5' ou 3' du locus BX-C [[Bantignies et al., 2011](#)]. Une limite aux propagations des interactions est donc observée aux bordures du locus *Bithorax*, bordures coïncidant à une frontière entre euchromatine et hétérochromatine. Une correspondance entre domaine structural et domaine fonctionnel existe donc au niveau de ce locus.

Les données Hi-C montrent également l'existence de domaines structuraux. L'analyse des matrices d'interactions met en évidence que la totalité du génome de l'homme et de la souris peut-être divisée en deux compartiments A et B [[Lieberman-Aiden et al., 2009](#); [Zhang et al., 2012](#)]. Ces segments de type A ou B, qui correspondent respectivement à de la chromatine ouverte (sensible à la DNase) ou fermée, interagissent avec leurs homologues ($A \leftrightarrow A$ et $B \leftrightarrow B$) sans cross-contacts

1.2- Des repliements contrôlés et dynamiques

(A \leftrightarrow B). D'autres données, obtenues par une méthode proche du Hi-C, aboutissent également à une segmentation en deux compartiments isolés l'un de l'autre [Kalhor et al., 2012]. A l'échelle du méga-base, les probabilités d'interactions sont compatibles avec une organisation en globule fractal (voir page 32) relativement individualisé par rapport à ses voisins [Lieberman-Aiden et al., 2009; Zhang et al., 2012]). Cette analyse consolide l'observation de la ségrégation spatiale des domaines fonctionnels.

La formation de domaines structuraux (appelés "globules chromatinien") a également été mise en évidence grâce à la méthodologie 5C [Baù et al., 2011]. En effet, les auteurs de cette étude ont analysé 750 interactions sur 500 kb contenant le locus α -globine dans deux types cellulaires : Gm12878 (lymphoblastoïde) et K562 (myéloïde à propriété hématocytaire). En couplant ces mesures d'interactions 5C avec de la modélisation de structure 3D, ils ont montré que cette région de 500 kb s'organise en un seul globule chez les cellules Gm12878 et en deux globules chez les cellules K562. Les gènes actifs sont plutôt retrouvés au centre de ces globules et les quelques gènes inactifs plutôt en périphérie. L'existence de globules laisse envisager un repliement de la chromatine qui autorise la formation d'unités structurales.

La FISH et les expériences de la famille 3C vont dans le sens d'une organisation de la chromatine en unités structurales de l'ordre de la mégabase, dont les frontières tendent à coïncider avec celles des marques fonctionnels (transcription, épigénétique). Le repliement à grande échelle de la chromatine, qui définit des domaines structuraux et fonctionnels, intègre donc nécessairement la contribution de plusieurs gènes ou régions régulatrices.

Nous reviendrons sur la segmentation structurale du génome telle que visualisée en Hi-C dans le chapitre suivant en montrant qu'elle peut également être reliée à la réplication. En outre, notre travail permettra l'identification de domaines structuraux dont les frontières coïncident avec des régions répliquées en début de phase S.

c- Maintien des frontières par CTCF et les cohésines

CTCF Cette protéine à doigt de zinc est largement impliquée dans l'organisation à grande échelle de la chromatine, et dans l'établissement des contacts chromatinien précédemment décrits. CTCF (CCCTC-binding factor) est par exemple impliqué dans la formation de "l'active chromatin hub" au niveau du locus bêta-globine [Splinter et al., 2006], mais aussi dans l'établissement de la conformation 3D du locus H19/Igf2 [Kurukuti et al., 2006] ou du locus apolipoprotéine [Mishiro et al., 2009].

En plus d'être impliqué dans des contacts entre portions de chromatine, CTCF est retrouvé à la jonction entre domaines chromatinien. En effet, un enrichissement en CTCF est observé au niveau des jonctions entre les domaines interagissant avec la lamina nucléaire et ceux n'interagissant pas [Guellen et al., 2008], ainsi qu'au niveau des jonctions entre les domaines riches ou pauvres en H3K27me3 [Cuddapah et al., 2009]. De plus, une étude basée sur un ChIA-PET avec un anticorps dirigé contre CTCF montre qu'il existe 4 types de configurations chromatinien ancrées par CTCF : (1) une boucle de chromatine ouverte ancrée par CTCF et ainsi isolée des régions voisines répressives, (2) le symétrique de la première situation, avec cette fois-ci une boucle répressive ancrée par CTCF

l'isolant des voisins actifs, (3) des interactions entre promoteurs et enhancers ancrées par CTCF, et (4) une boucle de chromatine ancrée par CTCF et située à la jonction entre un domaine actif et un domaine répressif [Handoko et al., 2011]. Cette dernière catégorie étant la plus abondante, elle illustre bien la capacité de CTCF à résider à la jonction de domaines fonctionnels, et conforte sa qualité d'organisateur du génome.

Le complexe cohésine Il est constitué de plusieurs protéines (entre autre Smc1, Smc3, Rad21 et SA1/2) et adopte une structure en anneau. Si le complexe cohésine est impliqué de longue date dans la cohésion des chromatides sœurs de la phase S jusqu'à l'anaphase, la manière dont s'effectue ce lien est encore débattu et deux modèles principaux s'opposent. Dans le premier, un anneau de cohésine suffirait pour enlancer les deux chromatides sœurs, alors que le deuxième implique deux complexes cohésines qui enlacent chacun une seule chromatide, les deux complexes étant reliés soit directement, soit *via* un facteur tierce (pour une revue, voir Carretero et al. [2010]).

Les complexes cohésines ont, en plus, un rôle à jouer dans la régulation de la transcription des gènes. Une maladie génétique, causée par une mutation dans un des facteurs responsables du recrutement des cohésines sur la chromatine, s'accompagne de profonds changements dans l'expression de nombreux gènes. Les gènes présentant un niveau d'expression altéré possèdent des sites de fixation aux cohésines au niveau de leur promoteur, alors que les gènes sans changement d'expression ne montrent pas une aussi forte fixation des cohésines à leur promoteur [Liu et al., 2009].

De nombreux sites de fixation des complexes cohésines correspondent également à des sites de fixation de CTCF [Kagey et al., 2010]. Les complexes cohésines ont donc été impliqués dans l'expression et la mise en place de la configuration 3D des loci pour lesquels CTCF est impliqué, à savoir le locus Igf2/H19 [Nativio et al., 2009], apolipoprotéine [Mishiro et al., 2009] et bêta-globine [Hou et al., 2010].

CTCF et les cohésines sont présents aux jonctions des domaines structuraux et joueraient un rôle dans la spécification de ces frontières. Dans ce manuscrit, nous serons amenés à discuter le positionnement de CTCF en lien avec la répliation.

1.2.4- Repliement et établissement de contacts entre chromosomes

Nous avons déjà abordé l'existence de territoires chromosomiques (voir page 33). Dans la description qui avait été faite, nous avons indiqué que l'ensemble de la chromatine d'un chromosome se retrouvait confiné à un territoire discret dans le noyau. Dans cette partie, cette discussion va être étendue : nous verrons que certaines portions de chromatine sont capables de sortir de leur territoire chromosomique de rattachement, permettant ainsi l'établissement de contacts avec d'autres chromosomes (*trans*-contacts). Nous ferons appels à ces notions dans la partie résultat de ce manuscrit (pages 112, 137 et 158).

a- Éviction du territoire chromosomique

Si la majorité des loci d'un chromosome est localisée au sein du territoire chromosomique, certaines régions sont capables d'en sortir. La plupart du temps, ces évictions du territoire chromosomique sont corrélées avec une forte expression ou avec l'activation du locus en question (pour des revues récentes, voir [Heard and Bickmore \[2007\]](#); [Fraser \[2006\]](#)).

Alors que le locus d'histocompatibilité majeur de classe II (chromosome 6) sort du territoire chromosomique dans seulement 10% des fibroblastes, lorsque son expression est induite par un traitement à l'interféron gamma, le taux d'éviction grimpe à 25%. Dans les lymphoblastoïdes B qui l'expriment constitutivement, ce locus est retrouvé en dehors du territoire du chromosome 6 dans 30% des cellules. Avec cette éviction, le locus d'histocompatibilité majeur de classe II colocalise plus souvent avec les territoires d'autres chromosomes (chr 1, 2 et 9) [[Branco and Pombo, 2006](#)].

De la même façon, un regroupement de 40 gènes impliqués dans la différenciation des kératinocytes est extrudé du territoire du chromosome 1 dans les kératinocytes [[Williams et al., 2002](#)]. Le locus Hoxb du chromosome 11 de la souris subit également une extrusion du territoire chromosomique lors de son activation [[Chambeyron and Bickmore, 2004](#); [Chambeyron et al., 2005](#)]. Enfin, un ensemble de gènes constitutivement exprimés du chromosome 11 humain montre aussi fréquemment une localisation en dehors de son territoire de rattachement [[Mahy et al., 2002](#)].

Dans ces 4 exemples, l'éviction est directement reliée à l'activation du locus ou bien à sa forte transcription. L'inactivation de l'ARN polymérase par l'actinomycine D diminue d'ailleurs la fréquence de ces localisations en dehors des territoires chromosomiques [[Mahy et al., 2002](#)]. Néanmoins, le niveau de transcription ne peut pas expliquer à lui seul la localisation extra-chromosomique pour plusieurs raisons : (1) Hoxd, qui s'exprime à des niveaux similaires dans les bourgeons caudal et membranaire, est seulement extrudé du chromosome 2 au niveau du bourgeon caudal [[Morey et al., 2007](#)]; (2) l'insertion d'un transgène Hoxb1 dans le locus Hoxd peut induire son éviction du territoire chromosomique en absence de toute transcription du transgène [[Morey et al., 2008](#)]; et (3) le locus alpha-globine humain se localise fréquemment en dehors de son territoire chromosomique, y compris dans les cellules n'exprimant pas ce locus [[Brown et al., 2006](#)]. Ces derniers travaux suggèrent donc que la transcription et l'éviction peuvent être, au moins partiellement, découplées.

Certaines portions de chromatine sont retrouvées hors de leur territoire chromosomique de rattachement. Ces évictions sont le plus souvent associées à un fort niveau d'expression, bien que le lien entre éviction et transcription reste à démontrer et n'est probablement pas direct.

b- Formation de *trans*-contacts

Ces évictions de boucles de chromatine en dehors de leur territoire chromosomique offre l'opportunité de former des contacts spécifiques entre deux loci portés par différents chromosomes (*trans*-interactions, à opposer aux *cis*-interactions au sein d'un même chromosome). Ainsi, un contact préférentiel a été observé entre le gène de l'interféron gamma (Infγ, une cytokine de la famille

des TH1, portée par le chromosome 10 murin) et le locus TH2 contenant les interleukines 5, 13 et 14 (chromosome 11 murin). Cette interaction, qui se met en place chez les lymphocytes T naïfs, devient beaucoup plus rare lorsque ces lymphocytes s'engagent dans une voie de différenciation plutôt TH1 (avec expression d'interféron gamma) ou TH2 (avec expression d'IL5, 13 et 14) [Spilianakis et al., 2005]. Cette interaction physique semblerait avoir une importance fonctionnelle. En effet, lorsque un site hypersensible à la DNase du locus TH2 (chr 11) est enlevé, ceci retarde l'expression de l'interféron gamma (chr 10) lors de l'activation des lymphocytes [Spilianakis et al., 2005]. Ce résultat indiquerait donc que les sites hypersensibles du chromosome 11 affectent la transcription de l'Infg localisé sur l'autre chromosome, vraisemblablement grâce à l'établissement de ce *trans*-contact.

De nombreux autres *trans*-contacts ont été décrits [Williams et al., 2010].

- Une association des loci alpha- et bêta-globine a été détectée par 3C utilisant des PCR nichées [Osborne et al., 2004] ainsi que par FISH [Osborne et al., 2004; Brown et al., 2006, 2008]
- Une association temporaire entre les deux centres d'inactivation du chromosome X survient juste avant l'inactivation d'un des deux chromosomes [Bacher et al., 2006; Xu et al., 2006a].
- Une association a été décrite entre les loci Igf2/H19 (chr 7) et Wsb1/Nf1 (chr 11) [Ling et al., 2006]. La FISH montre de plus que seule une paire d'allèle colocalise dans le noyau, et que cette colocalisation dépend de la région contrôle de l'empreinte parentale (ICR) du locus Igf2/H19 maternel. Cette interaction pourrait également avoir une importance fonctionnelle puisqu'une légère sur-expression de Nf1 et Wsb1 est observée lorsque une délétion de l'ICR maternelle d'Igf2/H19 est effectuée (en comparaison à une délétion de l'ICR paternelle).
- Le locus c-myc (chr 15, souris), dont l'expression est induite dans les lymphocytes B stimulés par l'IL4 et un anti-CD40, interagit plus fréquemment avec le locus Igh (chr 12, souris) après cette stimulation [Osborne et al., 2007].

c- Approches plus globales

Les différents exemples d'extrusion du territoire chromosomique et d'établissement de *trans*-contacts évoqués précédemment corrélerent avec des loci transcrits et régulés. Une manière de déterminer si cette localisation en périphérie, ou en dehors du territoire chromosomique est une caractéristique commune à toutes les régions transcrites, serait d'analyser les capacités d'interactions / la localisation de l'ensemble des loci exprimées.

Pour cela, un mélange de plus de 100 000 oligonucléotides fluorescents ciblant l'ensemble des exons transcrits du chromosome 2 murin a été utilisé comme sonde pour visualiser le territoire nucléaire occupé par tous les exons ("l'exome") du chr 2 murin. Le positionnement nucléaire de l'exome a ensuite été comparé à celui du territoire chromosomique de rattachement. Les observations microscopiques montrent que l'exome occupe un volume plus gros que le territoire chromosomique correspondant, en accord avec l'idée que les régions transcrites sont plus décondensées. Plus important, l'exome ne se superpose pas intégralement avec son territoire chromosomique et adopte une position plus centrale dans le noyau [Boyle et al., 2011]. Cette observation démontre donc que le positionnement en bordure, voire en dehors du territoire chromosomique est une propriété partagée par de nombreuses régions transcrites.

1.2- Des repliements contrôlés et dynamiques

Les données de TCC (Tethered Conformation Capture) sont obtenues par une méthode proche du Hi-C à ceci près qu'elle implique une ligation des fragments cross-linkés sur un support solide, ce qui, selon les auteurs, diminue le bruit de fond dû à des ligations aléatoires et permet ainsi une meilleure description des interactions entre-chromosomes [Kalhor et al., 2012]. Grâce à ces données, un index quantifiant la capacité d'un locus à établir des contacts avec d'autres chromosomes a été proposé. La pertinence et la définition de cet index sont largement soutenues par l'impressionnante corrélation observée entre la valeur de cet index et le positionnement vis-à-vis du territoire chromosomique (un index élevé étant associé à une extrusion du territoire chromosomique). Les valeurs de cet index ont donc pu être comparées sur l'ensemble du génome, démontrant que les régions de chromatine active établissent plus d'interaction avec les autres chromosomes que les régions de chromatine inactive. Cette étude confirme la propriété commune aux régions actives d'établir des *trans*-contacts, et donc de sortir de leur territoire chromosomique.

Des contacts entre différents chromosomes peuvent s'établir. Les régions de chromatine active peuvent être extrudées du territoire chromosomique. On verra dans la partie résultat que des contacts entre-chromosomes peuvent être observés en lien avec la réplication.

Conclusion

L'organisation à grande échelle de la chromatine permet la formation de contacts privilégiés, par exemple entre régions régulatrices ou entre régions corégulées. Ces contacts à long distance tendent à séparer la chromatine en deux groupes qui se ségrègent dans le noyau : un groupe actif qui interagit avec leurs semblables, et un groupe inactif qui interagit avec d'autres régions inactives. Les expériences de FISH combinées au développement des méthodologies de la famille 3C ont joué un rôle essentiel dans l'émergence de ces concepts dans les dernières années.

Les régions transcriptionnellement actives sont plutôt répliquées en début de phase S alors que les régions inactives sont plutôt répliquées en fin de phase S. En accord avec la ségrégation des régions actives et inactives dans le noyau, on montrera ultérieurement que les régions précoces et tardives occupent également des territoires distincts dans le noyau.

Dans la section suivante, nous verrons que ces contacts à longue distance peuvent s'effectuer au niveau de certains sous-compartiments du noyau.

1.3- Structure nucléaire et localisation des contacts

Le noyau contient un certain nombre de sous-compartiments aux fonctions plus ou moins connues. Le repliement non-aléatoire de la chromatine pourra favoriser la localisation de certaines séquences au sein de ces sous-compartiments.

1.3.1- Structure du noyau

Différents sous-compartiments existent dans le noyau (figure 14) [Spector, 2006; Zhao et al., 2009; Mao et al., 2011]

- Le nucléole correspond au lieu de transcription des ARNr. Les nucléoles d'une cellule sont formés par le regroupement des régions organisatrices du nucléole (NOR) qui contiennent des répétitions en tandem des gènes ribosomiaux. Chez l'homme, ces NOR sont trouvées sur les bras courts des chromosomes acrocentriques (chr 13, 14, 15, 21 et 22).
- Les corps de Cajal ont un diamètre de $\approx 0.5 \mu\text{m}$ et correspondent à des corps nucléaires impliqués dans l'assemblage et la maturation des petites particules ribonucléo-protéiques nucléaires (snRNP). Parmi les snRNP, certaines sont impliquées dans l'épissage des introns.
- La lamine nucléaire est située juste sous-l'enveloppe. Les lamines, qui sont des filaments intermédiaires, sont divisées en deux classes : la classe A correspondant aux lamines A et C qui dérivent du même gène LMNA par épissage alternatif, et la classe B correspondant aux lamines B1 et B2 qui dérivent des deux gènes LMNB1 et LMNB2. D'autres protéines structurales de la lamine nucléaire ont été décrites dont celles dont nous reparlerons : les protéines Lap (lamina associated polypeptide) et Emerin.
- Les corps d'épissage ("speckles") correspondent à des corps nucléaires enrichis en facteurs d'épissage. Ces corps nucléaires sont visualisés grâce à la protéine SC35.
- Les corps PML (ProMyelocytic Leukaemia) correspondent à des corps inframicroscopiques dont les rôles ne sont pas encore bien clairs. Ils seraient impliqués dans la régulation de la transcription, mais aussi dans la réparation de l'ADN et la réponse anti-virale (pour une revue, voir Bernardi and Pandolfi [2007]).
- Les usines de transcription correspondent à des accumulations locales d'ARN polymérase II.
- D'autres corps nucléaires ont également été décrits, notamment les corps OPT et les paraspeckles, mais nous n'en reparlerons pas par la suite. Par contre, en abordant plus tard les liens entre organisation de la chromatine et réplication, nous évoquerons un autre sous-compartiment nucléaire, les foyers à réplication, qui correspondent à des sites discrets où s'effectue la réplication (voir page 82 et suivantes)

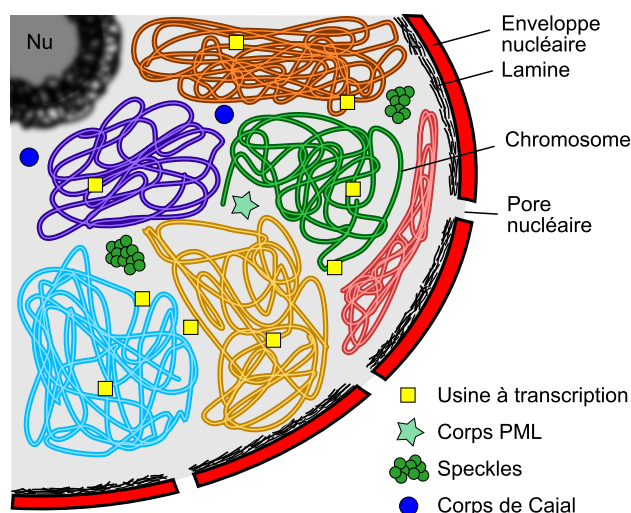


FIGURE 14 – Structure du noyau. Sur ce schéma sont représentés différents sous-compartiments du noyau : le nucléole (Nu), les usines de transcription, les corps PML, les speckles et les corps de Cajal. Le noyau est délimité par l'enveloppe nucléaire qui est ponctuée de pores. Juste sous l'enveloppe se trouve la couche de lamine.

1.3.2- Localisation à un compartiment actif

a- Les usines de transcription

Dans le noyau, la transcription a lieu au niveau de sites discrets visualisés par l'incorporation d'analogues d'UTP. Ces sites discrets correspondent à des accumulations locales d'ARN-polymérase II et sont appelés usine de transcription [Cook, 1999, 2002; Sutherland and Bickmore, 2009]. Ces usines de transcription, qui seraient plus de 1000 [Iborra et al., 1996], contiendraient l'ensemble de la machinerie moléculaire requise pour l'expression efficace des gènes actifs, dans un modèle où l'ADN passerait à travers l'usine contenant des polymérases immobiles [Sutherland and Bickmore, 2009; Schoenfelder et al., 2010a].

Les usines de transcription étant moins nombreuses que le nombre de gènes actifs d'une cellule, il a été proposé que chacune d'entre-elles contiendrait plusieurs gènes exprimés. Ainsi, l'association entre les loci alpha- et bêta-globine (évoquée page 49) pourrait s'effectuer au sein de la même usine de transcription [Osborne et al., 2004; Schoenfelder et al., 2010b]. De même, Myc et Igh pourraient partager la même usine de transcription [Osborne et al., 2007] (figure 15). Cette colocalisation entre gènes actifs au sein d'une même usine de transcription serait un phénomène fréquent puisqu'en combinant une analyse 4C et un ChIP avec un anticorps dirigé contre l'ARN polymérase phosphorylée (Sérine 5), il a été montré que 90% des loci qui interagissent avec la bêta-globine au sein d'une usine de transcription contiennent au moins un gène actif à proximité. De même, 80% des loci interagissant avec l'alpha-globine au sein d'une usine de transcription correspondent à des régions transcriptionnellement actives [Schoenfelder et al., 2010b].

b- Les corps d'épissage ou speckles

Dans le paragraphe précédent, nous avons dit que l'alpha- et la bêta-globine colocalisent au sein de la même usine de transcription [Osborne et al., 2004; Schoenfelder et al., 2010b]. Pourtant, d'autres groupes notent plutôt une colocalisation au niveau des corps d'épissage (speckles) [Brown et al., 2006, 2008] (figure 15). Leur ar-

gument est double. D'une part, ils observent une colocalisation avec SC35 qui est une protéine structurale de ce corps nucléaire. D'autre part, la distance moyenne mesurée entre les gènes α - et β -globine est de l'ordre de 500 nm, ce qui est beaucoup plus que la taille d'une usine de transcription (45-100 nm [Sutherland and Bickmore, 2009]), mais compatible avec la taille des speckles (0.8-1.8 μ m de diamètre [Spector, 2006]). L'implication de l'un ou l'autre corps nucléaire n'est pour l'heure pas tranchée.

SLCA1 (protéine membranaire bande 3 des érythrocytes) est également retrouvée à proximité des corps d'épissage (speckles) et les associations alpha-globine \leftrightarrow bêta-globine, SLCA1 \leftrightarrow alpha-globine, SLCA1 \leftrightarrow bêta-globine ont tendance à se retrouver en bordure du même corps d'épissage [Brown et al., 2006, 2008].

Comme les usines de transcription, les speckles renferment la forme phosphorylée de l'ARN polymérase II (pour une revue, voir Spector and Lamond [2011]) et de nombreux gènes / segments chromosomiques actifs montrent une localisation préférentielle à proximité des speckles. Par exemple, les bandes chromosomiques R (riches en gènes) contactent plus fréquemment les speckles que les bandes G, et ceci serait dû à la contribution de quelques gènes interagissant particulièrement avec ces corps d'épissage [Shopland et al., 2002]. De même, certains gènes essentiels pour la myogenèse (MyoD et cMyHC) se retrouvent à proximité des speckles dans les cellules où ils sont exprimés, mais pas dans les fibroblastes ou des myoblastes avant leur différenciation [Moen et al., 2004].

c- Les corps PML

La fonction des corps PML, notamment son lien avec la transcription, reste encore à clairement établir. Comme discuté par Bernardi and Pandolfi [2007], les corps PML contiennent à la fois des facteurs stimulant et réprimant la transcription. Néanmoins, il a pu être établi que certains loci se localisent au niveau des corps PML.

Ainsi, au moins un allèle du gène p53 colocalise avec les corps PML dans les cellules Jurkat (figure 15) [Sun et al., 2003]. D'autre part, de nombreuses mesures de distance entre différents loci et les corps PML attestent que les régions denses en gènes exprimés sont plus proches d'un corps PML que les régions inactives [Wang et al., 2004a], indiquant une association préférentielle entre chromatine active et corps PML.

Le corps PML pourrait jouer un rôle direct au niveau de l'expression génique en modulant les points d'ancrage de loci avec la matrice nucléaire. En effet, le locus du complexe majeur d'histocompatibilité (CMH) s'organise en boucles grâce à de nombreux points d'ancrage à la matrice nucléaire qui sont contrôlés par SATB1 (une des protéines de la matrice) et par PML [Kumar et al., 2007]. Étant donné que le knock-down de chacune des deux protéines altère à la fois l'expression du locus CMH et son ancrage à la matrice, ceci suggère que les corps PML pourraient moduler l'expression génique par la formation de boucle de chromatine. SATB1 et PML interagissant *in-vitro* et *in-vivo*, ce mode de régulation pourrait être fréquent.

Les loci exprimés et *a fortiori* les contacts entre ceux-ci, s'effectuent au sein de sous-compartiments du noyau. L'implication des usines de transcription ou des corps d'épissage (speckles) est néanmoins encore débattue.

1.3- Structure nucléaire et localisation des contacts

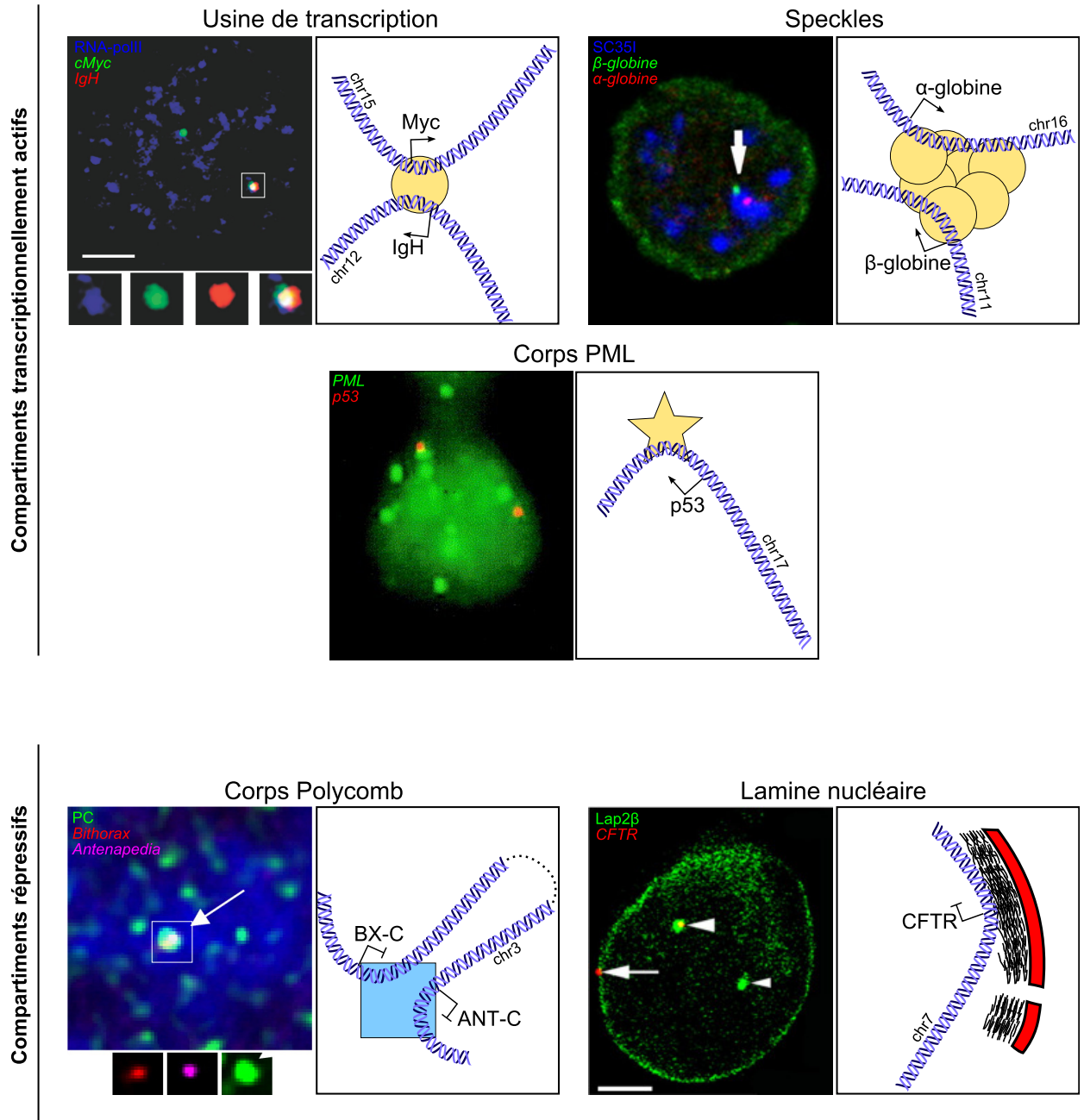


FIGURE 15 – Localisation de loci à différents sous-compartiments nucléaires. Myc et IgH peuvent se localiser à la même unité de transcription lorsqu'ils sont exprimés. L'image d'immuno-FISH montre en bleu l'ARN-polII phosphorylée, en rouge IgH et en vert Myc dans des lymphocytes B stimulés murins. Barre = 2µm [Osborne et al., 2007]. L'alpha- et la bêta-globine peuvent s'associer au sein du même speckle. L'image d'immuno-FISH montre en bleu SC35, en rouge l'alpha-globine et en vert la bêta-globine dans des érythroblastes intermédiaires humains [Brown et al., 2006]. Un des deux allèles de p53 s'ancre préférentiellement au niveau des corps PML dans les cellules Jurkat (vert = PML, rouge = p53) [Sun et al., 2003]. Au niveau de la tête d'embryon de drosophile, les complexes Antennapædia (violet) et Bithorax (rouge) du chromosome 3 se retrouvent au niveau du même corps Polycomb (vert) [Bantignies et al., 2011]. Enfin, CFTR est localisé au niveau de l'enveloppe nucléaire dans les cellules de neuroblastome où il n'est pas exprimé (vert = Lap2β, rouge = CFTR; la tête de flèche indique un des allèles CFTR localisé à une invagination de l'enveloppe nucléaire) [Zink et al., 2004].

Reproduit avec autorisation de Osborne et al. [2007], de Brown et al. [2006] originellement publié dans *J Cell Biol* doi:10.1083/jcb.200507073, de Sun et al. [2003] Copyright 2003 avec l'autorisation d'Elsevier, de Bantignies et al. [2011] Copyright 2011 avec l'autorisation d'Elsevier et de citet-Zink2004 originellement publié dans *J Cell Biol* doi:10.1083/jcb.200404107.

1.3.3- Localisation à un compartiment répressif

Certains loci réprimés sont quant à eux retrouvés au sein de compartiments répressifs du noyau.

a- La lamine nucléaire

De nombreuses expériences démontrent que la périphérie nucléaire correspond à un compartiment répressif.

L'analyse d'images obtenues par microscopie électronique à transmission montrent une plus forte opacité aux électrons en périphérie du noyau, interprétée comme une plus forte compaction de la chromatine à proximité de la lamine nucléaire. Une expérience de Dam-ID, avec l'expression de la lamine B1 fusionnée à une méthylase de l'ADN, confirme que la lamine correspond à un environnement répressif [Guelen et al., 2008]. En effet, les portions du génome qui deviennent méthylées et qui sont donc associées avec la lamine nucléaire sont pauvres en gènes exprimés, dépourvues d'ARN polymérase II et possèdent la signature épigénétique d'une chromatine répressive (fort H3K27me3 et faible H3K4me2).

Certains loci quittent l'enveloppe nucléaire lorsque qu'ils sont activés. En effet, les loci IgH (chaîne lourde d'immunoglobuline) et Ig κ (chaîne légère kappa d'immunoglobuline) sont localisés à la périphérie nucléaire dans les cellules n'appartenant pas à la lignée B, y compris dans les cellules ES, ce qui suggère que la localisation à la lamine nucléaire correspondrait à l'état par défaut [Kosak et al., 2002]. *A contrario*, dans les cellules de la lignée B où ces loci sont exprimés, un positionnement plus central est observé. Cette différence de localisation nucléaire selon le niveau d'expression n'est pas un cas isolée. Le gène CFTR ainsi que les 2 gènes qui l'entourent (GASZ et CORTBP2) sont également localisés contre la lamine nucléaire dans les cellules où ils ne sont pas exprimés (figure 15), mais adoptent une position plus centrale dans le noyau lorsque les gènes sont actifs [Zink et al., 2004].

Pour démontrer le lien de causalité entre le positionnement en périphérie du noyau et l'établissement d'un contexte chromatinien répressif, des expériences d'ancrage artificiel à la lamine nucléaire ont été conduites. La stratégie consiste à utiliser un transgène intégré dans le génome contenant des répétitions en tandem de l'opérateur lac (lacO), et de l'ancrer à la lamine nucléaire par l'expression d'une protéine fusion contenant lacI et une protéine de la matrice nucléaire (lamine B1 [Kumaran and Spector, 2008], Lap2 β [Finlan et al., 2008] ou un fragment d'Emerin [Reddy et al., 2008]). Grâce à ces stratégies, il a été démontré que l'adressage artificiel à la lamine nucléaire conduit à une diminution d'expression du transgène associée à une hypo-acétylation du locus [Reddy et al., 2008; Finlan et al., 2008]. Plus intéressant, cet adressage à la lamine nucléaire, en plus d'induire une diminution d'expression du transgène, peut entraîner la répression de certains gènes alentours [Finlan et al., 2008], démontrant ainsi l'importance du micro-environnement comme niveau de régulation de la transcription.

b- Les corps Polycomb

Chez la drosophile, le complexe Antennapedia et le complexe Bithorax, qui sont tous les deux régulés par les facteurs Polycomb, interagissent ensemble dans les tissus où les deux loci sont réprimés [Bantignies et al., 2011]. L'analyse de cette

1.3- Structure nucléaire et localisation des contacts

colocalisation par hybridation *in-situ* montre que ces deux segments, qui sont séparés de 10 Mb, interagissent au niveau du même corps Polycomb (figure 15). Cette association est perdue chez des mutants déficients en Polycomb, indiquant l'importance de ce complexe répresseur pour l'établissement du contact. D'autre part, grâce à des approches 4C ou Hi-C, il a été montré que l'ensemble des loci réprimés par les complexes Polycomb manifestent une interaction préférentielle [Bantignies et al., 2011; Tolhuis et al., 2011; Sexton et al., 2012]. Il reste néanmoins à établir si ce regroupement des régions régulées par les complexes Polycomb revêt une importance fonctionnelle pour l'établissement de la répression.

c- Périphérie du nucléole

Le nucléole est un exemple flagrant d'interactions entre chromosomes, puisque les gènes ribosomiaux (≈ 400 copies chez l'homme), localisés sur 5 chromosomes différents (chr 13, 14, 15, 21 et 22), s'agrègent à la fin de la mitose pour former le/les nucléole(s).

Si le nucléole en tant que tel correspond à un environnement transcriptionnellement actif (synthèse des gènes ribosomiaux), la microscopie électronique montre que les pourtours du nucléole sont enrichis en hétérochromatine constitutive. Pour confirmer cette observation histologique, des nucléoles ont été isolés et l'ADN associé (l'ADN du nucléole et l'ADN péri-nucléolaire) a été analysé par hybridation *in-situ* et séquençage haut-débit de type 454 [Németh et al., 2010]. Cette étude montre que l'ADN péri-nucléolaire regroupe 4% du génome. Parmi les séquences identifiées, on trouve des séquences satellites, peu de SINE et des gènes réprimés. Ces séquences péri-nucléolaires correspondent à des portions de chromatine répressive marquées par H3K27me3, H3K9me3 et H4K20me3, confirmant la nature répressive de la chromatine péri-nucléolaire.

Gènes actifs et inactifs occupent des sous-compartiments nucléaires distincts. Les régions inactives sont retrouvées en périphérie du noyau, dans les corps Polycomb, ou aux pourtours du nucléole. A l'inverse, les régions actives occupent plutôt les usines de transcription, corps d'épissage et corps PML. Cette différence de localisation n'est pas une spécificité de la transcription puisque nous verrons dans le chapitre suivant que les régions répliquées en fin de phase S interagissent également plus avec la lamine nucléaire que les régions répliquées en début de phase S.

Conclusion

Dans ce chapitre, nous avons pu démontrer que la chromatine est organisée dans le noyau. On observe une hiérarchie de niveaux de repliement (nucléofilament, fibre de 30 nm, boucles géantes de chromatine, territoires chromosomiques), dont chacun atteste du lien étroit entre organisation tridimensionnelle et fonction. Nous avons en effet montré que des boucles locales (~ 100 kb) peuvent supporter la communication entre séquences régulatrices et gènes cibles (paragraphe 1.2.2), que les niveaux de repliements supérieurs (~ 1 Mb) supportent la formation de domaines structuraux dont les frontières coïncident avec celles de

domaines fonctionnels et qui ségrègent dans l'espace nucléaire en fonction de leur statut chromatinien (paragraphe 1.2.3). Ces domaines sont contenus dans leur territoire chromosomique (~100 Mb) même si des évictions sont possibles pour les loci qui sont exprimés (paragraphe 1.2.4). Deux complexes protéiques, CTCF et les cohésines, sont retrouvés aux jonctions des domaines nucléaires, et pourrait jouer un rôle dans leur spécification.

Le niveau d'organisation du noyau ne se limite pas à l'agencement non aléatoire de la chromatine. Le noyau contient en effet de nombreux sous-compartiments réalisant des fonctions nucléaires précises (paragraphe 1.3.1), dont certains correspondent à des micro-environnements actifs en terme de transcription (usines de transcription, corps d'épissage, corps PML), et d'autres à des micro-environnements plus répressifs (lamines nucléaires, pourtours du nucléole). Avec la matrice nucléaire (page 31), ils assurent l'ancrage et la colocalisation de certains loci.

Si l'accent a été placé sur les liens entre transcription et organisation nucléaire, la réplication se déroule également dans ce contexte de noyau organisé. Façonne-t-elle également le noyau ? Certains aspects de l'organisation nucléaire sont-ils reliés au déroulement de la réplication ? Nous répondrons à ces questions dans le chapitre suivant.

La réplication de l'ADN

2.1 Mécanismes généraux de la réplication de l'ADN	60
2.1.1 Déroulement de la réplication	60
<i>Activation de l'origine et ouverture de la double hélice</i>	60
<i>Polymérisation d'ADN</i>	62
2.1.2 Les origines de réplication identifiées chez l'homme par biologie moléculaire	63
2.1.3 Contrôle du choix des ORI et de leur activation coordonnée	64
<i>Sélection des origines</i>	64
<i>Etablissement du timing de réplication</i>	64
<i>Une seule activation des ORI par cycle</i>	66
2.2 Approches génomiques et à haut-débit pour l'étude de la réplication	66
2.2.1 Biais de composition en nucléotides et réplication	67
2.2.2 Identification de sites de liaison du complexe pré-réplicatif (pre-RC)	69
2.2.3 Détection de sites d'initiation	70
2.2.4 Établissement du timing de réplication	71
<i>Les différentes méthodes</i>	71
<i>Propriétés des régions précoces et des régions tardives</i>	72
<i>Conservation et modification du timing de réplication</i>	74
<i>Identification de domaines fonctionnels de timing</i>	75
2.3 Structuration de la chromatine associée à la réplication	78
2.3.1 Réplication et boucles d'ADN	79
<i>Halo nucléaire et réplication</i>	79
<i>Matrice et réplication – Correspondance structurale</i>	80
<i>Matrice et réplication – Correspondances fonctionnelles</i>	81
2.3.2 Les foyers de réplication	82
<i>La réplication au niveau de foyers discrets</i>	82
<i>Les foyers correspondent à des usines de réplication</i>	84
<i>Fonctionnement de l'usine de réplication</i>	85
2.3.3 Contacts chromatiniens et réplication	86
<i>Chez la levure – interactions entre ARS</i>	86
<i>Chez l'homme – couplage du timing de réplication et des données Hi-C</i>	87
<i>Limites de ces études chez l'homme</i>	88

Au cours du chapitre précédent, l'accent a été placé sur l'expression des gènes et sa relation avec l'organisation de la chromatine. Néanmoins, la conservation du patrimoine génétique au cours des divisions cellulaires nécessite une duplication de cette information. Cette duplication (ou réplication) s'effectue au cours de la phase S du cycle cellulaire et se déroule alors que la chromatine est, comme nous l'avons vu, organisée en boucles et domaines structuraux. Ce chapitre nous amènera à discuter les interconnexions entre réplication et organisation de la chromatine. Pour cela, nous suivrons le cheminement suivant :

- 1 nous détaillerons brièvement les mécanismes moléculaires permettant la duplication du matériel génétique afin de positionner les acteurs auxquels nous ferons appel dans la suite du chapitre ;

- 2 nous discuterons les différentes approches permettant l'étude de la réplication sur l'ensemble du génome (l'étude que nous proposerons dans la suite du manuscrit fera appel à ces méthodologies à haut-débit, il convient donc de les introduire dès maintenant) ;
- 3 et dans une dernière partie, nous discuterons abondamment les liens déjà établis entre organisation de la chromatine et réplication, en insistant sur les questions encore en suspens.

2.1- Mécanismes généraux de la réplication de l'ADN

Une duplication de l'ADN est indispensable pour transmettre, après division cellulaire, une copie complète de l'information génétique aux deux cellules filles. La réplication de l'ADN s'effectue au cours de la phase S (pour synthèse) du cycle cellulaire qui dure approximativement 6-8 h et est encadrée par les deux phases G1 et G2 (phases de croissance).

La réplication de l'ADN fait appel à une machinerie enzymatique bien connue. Puisque nous serons amenés à faire appel à ces acteurs enzymatiques dans la suite du manuscrit, nous allons les présenter succinctement au début de cette section.

2.1.1- Déroulement de la réplication

a- Activation de l'origine et ouverture de la double hélice

L'initiation correspond à la phase de démarrage de la réplication au niveau d'une origine de réplication. Ceci nécessite le recrutement de plusieurs complexes protéiques.

Les origines de réplication (ORI) et le complexe pré-répliatif Chez *E. coli*, il existe une et une seule origine de réplication, appelée *oriC*. D'une taille de 250 bp, elle contient des répétitions servant chacune de plateforme au recrutement de DnaA. L'oligomérisation de plusieurs DnaA facilite alors l'ouverture de la double hélice d'ADN au niveau d'une région adjacente riche en AT. Cette ouverture locale de la double hélice suffit à recruter la machinerie de réplication [Mott and Berger, 2007].

Chez *S. cerevisiae*, une origine de réplication est également spécifiée par sa séquence, par contre au lieu d'avoir une seule origine, on en compte approximativement 400. Appelées ARS (Autonomously Replicating Sequence) pour leur capacité à induire la réplication d'un plasmide, elles sont riches en AT, contiennent des motifs répétés et sont reconnues par les complexes ORC (Origin Recognition Complex). La liaison de ces complexes ORC, suivi du recrutement en phase G1 de cdc6, cdt1 et des hélicases Mcm (Mini Chromosome Maintenance) forment le complexe pré-répliatif (pre-RC). Son activation en début de phase S déclenchera le recrutement de l'ADN polymérase et autres facteurs nécessaires pour la réplication [Méchali, 2010].

2.1- Mécanismes généraux de la réplication de l'ADN

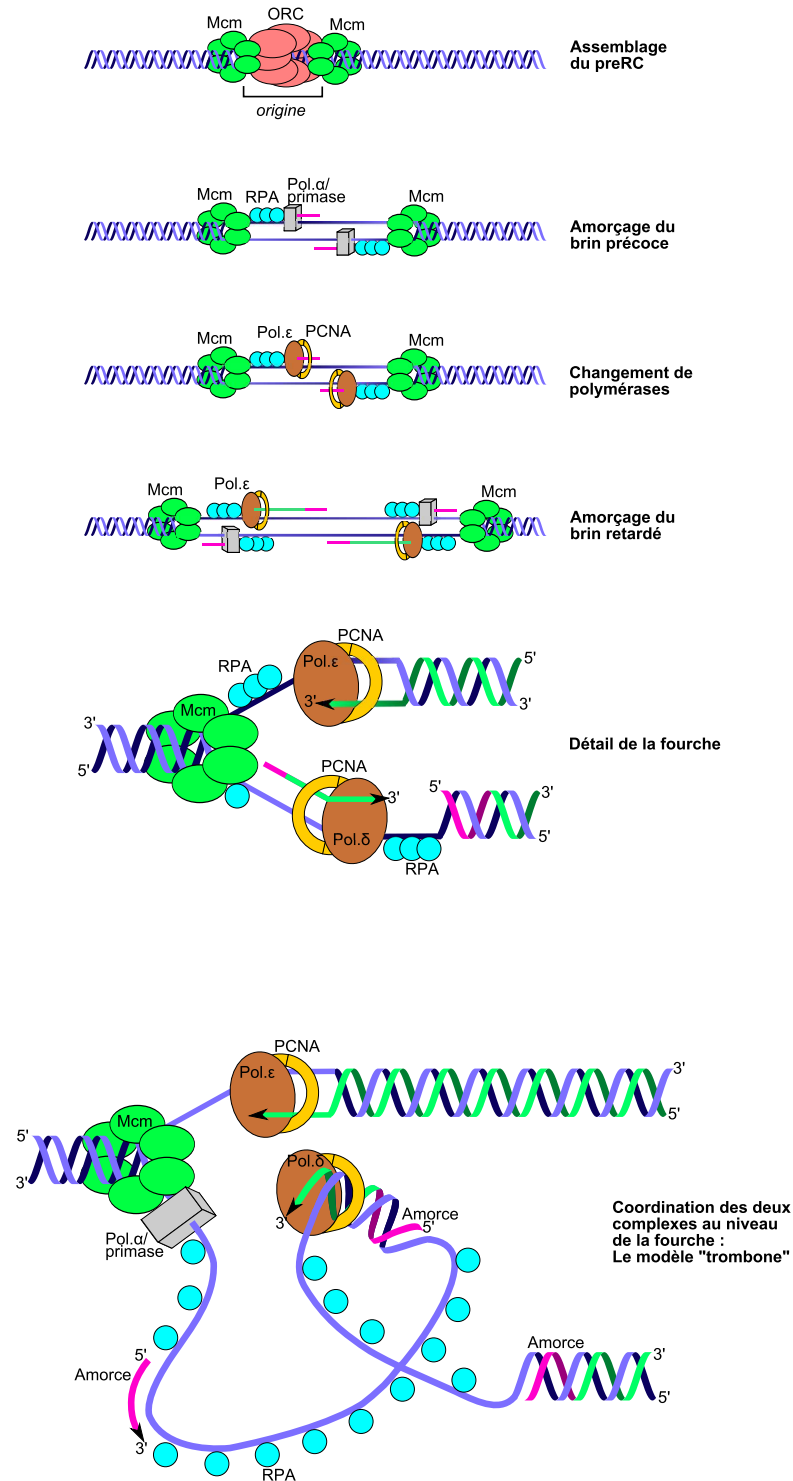


FIGURE 16 – La réplication de l'ADN chez les eucaryotes. De bas en haut : (1) Assemblage en phase G1 du pre-RC composé des complexes ORC et Mcm ; (2) Ouverture de la double hélice, recrutement des protéines RPA qui stabilisent l'ADN simple brin, et amorçage de la réplication par le complexe polymérase α / primase qui permet la synthèse de quelques nucléotides d'ARN suivis d'une trentaine de nucléotides ADN ; (3) Remplacement de la polymérase α par la polymérase ϵ avec son facteur de processivité PCNA ; (4) Le complexe polymérase α / primase est recyclé pour l'amorçage de la réplication sur le brin retardé ; (5) La fourche en cours d'élongation est détaillée avec notamment la polymérase ϵ sur le brin précoce et la polymérase δ sur le brin retardé ; (6) Pour coordonner les complexes du brin direct et du brin retardé, une organisation en "trombone" a été proposée, de telle sorte que la boucle formée permette une progression de la polymérase δ dans le même sens que la polymérase ϵ .

Chez les métazoaires, les origines de réplication ne sont pas spécifiées par la séquence [Méchali, 2010]. On estime qu'il y aurait au moins 30 000 origines de réplication chez l'homme.

Ouverture de la double hélice et activité primase Chez les bactéries, la partie N-terminale de DnaA est capable de recruter DnaB, une hélicase, au niveau de la zone désappariée. Ceci contribue à étendre la zone désappariée. L'ADN simple brin formé par ouverture de la double hélice est stabilisé et protégé par les protéines SSB (single-strand binding). Chez les eucaryotes, le principe est globalement le même. Les hélicases Mcm (Mini Chromosome Maintenance) recrutées en phase G1 sont activées en phase S, conduisant à une ouverture de la double hélice (figure 16). Cette région désappariée, qui est protégée et stabilisée par les protéines RPA (Replication Protein A), permet le recrutement de l'ADN-polymérase α accompagnée de sa primase.

b- Polymérisation d'ADN

Activité ADN polymérase et processivité Les ADN-polymérases sont incapables d'initier la synthèse d'ADN si une extrémité 3'-OH n'est pas disponible. Chez les bactéries, l'ouverture de la double hélice conduit au recrutement de DnaG qui possède une activité ARN-polymérase (primase). La courte séquence d'ARN (≈ 10 nt) synthétisée par DnaG servira ensuite d'amorce pour l'ADN polymérase III. Chez les eucaryotes, un processus similaire (mais en deux étapes) est observé : dans un premier temps, la polymérase α / primase permet la formation d'un court fragment d'ARN, lequel est ensuite prolongé par un petit fragment d'ADN de 30 nt synthétisé par la polymérase α . C'est à l'extrémité de ces 30 nt d'ADN que s'amorcera l'élongation grâce aux activités enzymatiques des polymérases ϵ et δ (figure 16).

Pour assurer la processivité des ADN-polymérases, le facteur β chez les bactéries ou PCNA (Proliferating cell nuclear antigen) chez les eucaryotes se logent juste à l'arrière des polymérases, en formant un anneau encerclant l'ADN.

Réplication semi-discontinue Étant anti-parallèles, les deux brins ne sont pas symétriques : l'un sera répliqué de manière continue (brin précoce), et l'autre de manière discontinue en impliquant la formation d'Okazaki (brin retardé). La synthèse de chaque fragment d'Okazaki impliquera donc une activité primase et une activité ADN-polymérase. Les fragments d'Okazaki ont une taille de 1000-2000 bp chez les procaryotes, et de 100-200 bp chez les eucaryotes. Les amorces de chaque fragment d'Okazaki seront remplacées par des ADN polymérases, avant qu'une ADN ligase ne relie les fragments d'Okazaki entre eux. Chez les eucaryotes, l'ADN-polymérase ϵ se charge de l'élongation sur le brin précoce et l'ADN-polymérase δ de l'élongation sur le brin retardé. Pour coordonner le complexe protéique se chargeant du brin précoce avec celui se chargeant du brin retardé, une structure en trombone a été imaginée (figure 16).

La duplication du matériel génétique fait appel à de nombreux acteurs protéiques dont différents complexes enzymatique (ARN-polymérase, ADN-polymérase, hélicase...). La biochimie de la réplication ne sera plus abordée dans ce manuscrit, néanmoins il nous arrivera de faire référence à certains acteurs (PCNA, ADN-polymérase, RPA...)

2.1.2- Les origines de réplication identifiées chez l'homme par biologie moléculaire

Chez les métazoaires, on appelle réplicon la séquence d'ADN contenue entre deux origines de réplication. Les réplicons humains ont une taille de 40 kb, comme ceci a pu être mesuré sur fibre d'ADN étirée après l'incorporation de nucléotides modifiés au niveau des zones en cours de réplication [Guilbaud et al., 2011]. Ramené à la taille du génome humain, cela représente plusieurs dizaines de milliers d'origines de réplication (ORI) pour l'ensemble du génome. Néanmoins, seule une quinzaine est réellement caractérisée par biologie moléculaire chez l'homme.

c-MYC Une origine de réplication a été identifiée approximativement 1,5 kb en amont du gène c-MYC, au niveau de son promoteur. Cette ORI a été positionnée par isolation de brins naissants [Vassilev and Johnson, 1990], ainsi que par étude de leur polarité le long du locus c-MYC [Waltz et al., 1996]. Néanmoins, si le promoteur de C-MYC semble correspondre au site d'initiation principale, une zone d'initiation secondaire a été identifiée au niveau de l'exon 2 du gène Tao et al. [2000].

Lamine B2 Une origine de réplication a également été identifiée en 3' du gène LMNB2. La zone d'initiation a d'abord été localisée dans une zone de quelques kb située entre les deux gènes LMNB2 et TIM13 (précédemment nommé ppv1) [Biamonti et al., 1992], avant d'être positionnée au nucléotide près à l'extrémité 3' du gène LMNB2 [Abdurashidova et al., 2000]. Un fragment de 1,2 kb contenant cette ORI et inséré en position ectopique dans le génome est toujours capable d'initier la réplication [Paixão et al., 2004], indiquant que l'activation de l'ORI de la lamine B2 ne nécessite pas de régulateurs plus distants.

β -Globine Cette ORI est localisée en 5' du gène β -globine et n'est active que dans les tissus exprimant le locus globine [Kitsberg et al., 1993]. Une région de 8 kb centrée sur cette origine et placée en position ectopique est capable d'initier la réplication, indiquant que ces 8 kb sont suffisant pour spécifier une origine de réplication [Aladjem et al., 1998]. Des délétions successives ont en outre montré que cette région contient deux modules capables d'initier indépendamment la réplication, dont l'un est localisé 2 kb en amont du gène β -globine, et l'autre comprend ses deux premiers exons [Wang et al., 2004b].

Les autres origines identifiées par biologie moléculaire chez l'homme
D'autres origines de réplication ont été décrites chez l'homme, y compris [Aladjem et al., 2006; Schepers and Papior, 2010] :

- dans la partie 5' du gène DNMT1 (DNA-méthyltransférase 1)
- en amont du gène HPRT
- au niveau du promoteur partagé par les deux gènes divergents MCM4 (Minichromosome maintenance complex component 4) et PRKDC (Protein kinase, DNA-activated, catalytic polypeptide)
- à proximité du gène RPE (Ribulose-5-phosphate 3-épipérase)
- au niveau du promoteur de HSP70 (Heat shock protein 70), TOP1 (DNA topoisomérase I), DNF4, FMR1 (Fragile X Mental Retardation protein 1)
- en aval du gène GM-CSF (Granulocyte Macrophage Colony-Stimulating Factor)
- au niveau du gène IFNA2 (Interféron alpha 2)

Seules une quinzaine d'origines de réplication est connue chez l'homme. Elles ne possèdent pas de séquences communes, mais la plupart d'entre elles se localisent à proximité des régions codantes [Aladjem et al., 2006]. Dans la partie résultat de ce manuscrit, l'organisation de la chromatine autour de l'origine c-myc sera analysée par 4C (page 152 et page 156).

2.1.3- Contrôle du choix des ORI et de leur activation coordonnée

a- Sélection des origines

Seule une fraction des sites de liaisons de protéines Mcm est effectivement convertie en origines de réplication. Les protéines Mcm pourraient en effet spécifier des origines de sauvegarde qui seraient activées en cas de problème lors de la réplication (pour une revue, voir Méchali [2010]). Il existe donc une étape décisionnelle où sont sélectionnées les ORI qui seront réellement activées en phase S.

Cette sélection s'effectue dans les quelques heures après le début de la phase G1. Lorsqu'un noyau de cellules CHO (Chinese Hamster Ovary) est incubé dans un extrait d'ovocyte de Xénope, ceci déclenche sa réplication. L'utilisation de ce système expérimental avec des noyaux en milieu de phase G1 (4 h après le relâchement d'un blocage en métaphase) conduit à l'activation de l'ORI du locus DHFR (dihydrofolate reductase) à la même position que celle initiant la réplication *in-vivo*. Maintenant, si la même expérience est réalisée, mais cette fois-ci avec des noyaux en début de phase G1 (1 h après la levée d'un blocage en métaphase), le noyau entre toujours en réplication, mais l'ORI utilisée pour répliquer le locus DHFR n'est plus la même. Les auteurs notent même des initiations aléatoires le long du locus étudié [Wu and Gilbert, 1996]. Cette observation n'est pas une particularité du locus DHFR. En effet, lorsqu'un noyau en début de phase G1 est incubé dans un extrait d'ovocyte de Xénope, la réplication *in-vitro* utilise de manière générale des ORI qui sont différentes de celles normalement utilisées *in-vivo*. Pour que le jeu d'ORI utilisées *in-vitro* et *in-vivo* concordent, il faut que l'expérience soit effectuée sur un noyau à une étape plus avancée de la phase G1 [Li et al., 2003]. Ces expériences indiquent donc qu'il existe 4 h après le début de la phase G1, une étape décisionnelle où sont sélectionnées les ORI qui seront activées en phase S.

La manière dont s'effectue cette sélection n'est pas vraiment connue. On sait néanmoins que ce choix est en partie stable et transmis d'une génération cellulaire à l'autre puisque l'analyse de fibres d'ADN étirées montre que l'incorporation de nucléotides modifiés au début de la phase S de cycles successifs peut s'effectuer au même endroit [Jackson and Pombo, 1998; Li et al., 2003].

b- Établissement du timing de réplication

Les ORI s'allument de manière coordonnée, certaines en début de phase S et d'autres en fin, en suivant un programme de réplication précis. Pour déterminer à quel moment de la phase G1 sont sélectionnées les loci qui se répliqueront précocement et ceux qui se répliqueront tardivement, des noyaux de cellules CHO à différentes étapes de la phase G1 sont incubés dans un extrait d'ovocyte de Xénope. Ceci déclenche leur réplication prématurée. L'ADN naissant peut alors

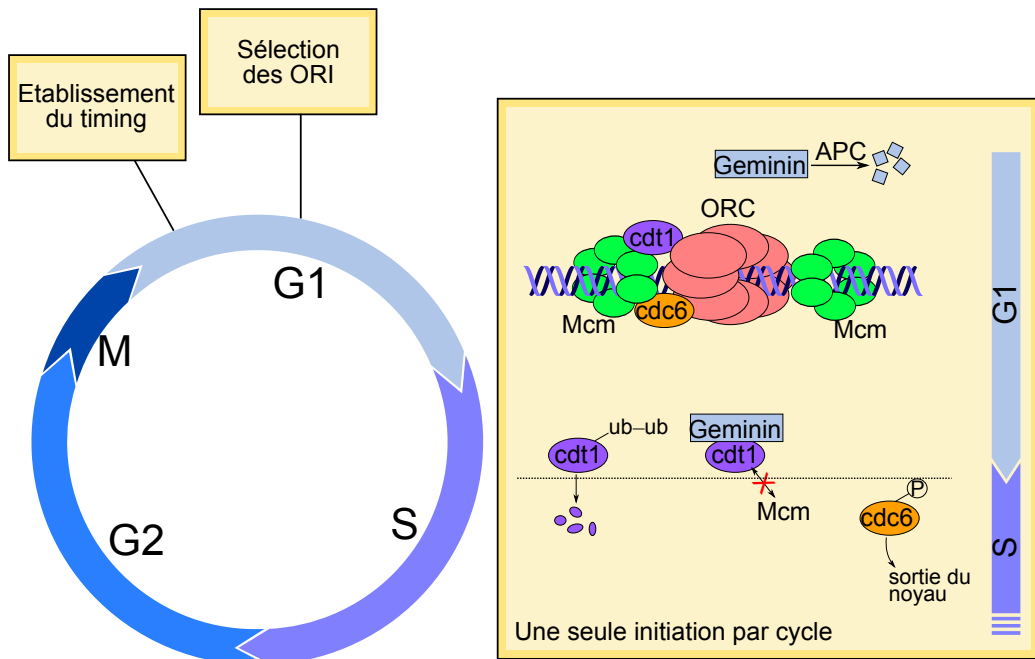


FIGURE 17 – Quelques points de régulation de la réplication. En début de phase G1 le timing de réplication est établi, puis plus tard, toujours en phase G1, les origines de réplication qui s'activeront sont sélectionnées. Pour assurer une seule activation des ORI par cycle, le niveau de cdt1 et de cdc6, deux composants du pre-RC, est régulé. En début de phase S, cdt1 est poly-ubiquitinylée, conduisant à sa dégradation par protéolyse. De plus, la Gémimine, qui en phase G1 était dégradée par le protéasome, séquestre cdt1 en phase S, l'empêchant ainsi d'interagir avec les protéines Mcm. En phase S, cdc6 est phosphorylée, ce qui entraîne sa sortie du noyau. Ensemble, ces mécanismes contribuent à empêcher la reformation du pre-RC, et donc les initiations supplémentaires.

être isolé dans les 30 premières minutes après le déclenchement de la phase S, ou 2 h après son démarrage. Celui-ci est ensuite radiomarqué, et hybridé sur des séquences connues pour se répliquer soit en début de phase S (locus DHFR), soit en fin de phase S (LINE C1 et C3). Lorsque cette expérience est réalisée avec des noyaux de fin de phase G1 (6 h après la levée du blocage en métaphase), l'hybridation montre que le locus DHFR se réplique avant les LINE C1 et C3, conformément à ce qu'il se passe *in-vivo*. Par contre, si cette expérience est réalisée avec des noyaux de début de phase G1 (1 h après la levée du blocage en métaphase), DHFR, C1 et C3 se répliquent simultanément. Ceci démontre que le timing de réplication est établi entre 1h et 6h après la levée d'un blocage en métaphase [Dimitrova and Gilbert, 1999].

Il est intéressant de remarquer que l'établissement du timing s'effectue avant la sélection des origines de réplication. En effet, alors que 2 h après la levée d'un blocage en métaphase, les ORI ne sont encore spécifiées, le timing de réplication semble déjà établi [Dimitrova and Gilbert, 1999; Li et al., 2003]. Ceci signifie que le programme temporel de la réplication est établi sans pour autant initier la réplication au niveau des ORI qui seraient utilisées *in-vivo*.

Plus surprenant, l'établissement du timing de réplication en début de phase S coïncide avec le rétablissement de l'organisation du noyau. En sortie de mitose, le noyau est désorganisé et quelques heures sont nécessaires pour rétablir l'hétérochromatine en périphérie du noyau et autour du nucléole. Lorsque cette réorganisation du noyau est achevée, le timing de réplication est aussi établi [Dimitrova and Gilbert, 1999; Li et al., 2001]. Ceci suggère que les deux phénomènes pourraient être connectés [Gilbert et al., 2010].

c- Une seule activation des ORI par cycle

Il est essentiel d'assurer une et une seule phase de réplication par cycle. Nous avons vu que la spécification des ORI potentielles s'effectue en phase G1 par la formation du pre-RC (pre-Replication Complex, voir page 60). La formation de ce pre-RC se déroule en plusieurs étapes : la liaison à l'ADN des complexes ORC, suivi du recrutement du complexe Mcm par les protéines cdt1 et cdc6.

Au démarrage de la phase S, le niveau de cdt1 et cdc6 disponible dans le noyau chute (par protéolyse et séquestration de cdt1, ainsi que par expulsion hors du noyau de cdc6) (figure 17). Or, ces deux protéines sont nécessaires pour assembler de nouveaux complexes pré-réplicatifs. La séquestration / destruction de cdc6 et cdt1 empêche donc la formation de nouveaux pre-RC après le démarrage de la phase S, ce qui permet d'éviter toute initiation additionnelle. La contribution de ces deux protéines dans l'assemblage du pre-RC et comment leur niveau est régulé pour maintenir une seule initiation par cycle cellulaire ont été récemment discutés par [Depamphilis and Blow \[2006\]](#) et [Truong and Wu \[2011\]](#).

L'initiation de la réplication est une étape particulièrement contrôlée. Le programme temporel de la réplication, ainsi que la sélection des ORI qui s'activeront sont établis en phase G1. La phase S démarrée, il sera impossible de démarre une nouvelle vague d'initiations avant la division cellulaire.

Conclusion

La réplication de l'ADN, qui s'effectue en phase S, requiert le fonctionnement coordonné de nombreuses enzymes. Elle débute au niveau d'origines de réplication (ORI) dont seule une quinzaine est bien caractérisée par biologie moléculaire chez l'homme. Ces ORI s'activent en suivant un programme temporel précis qui est établi en début de phase G1. Le démarrage de la réplication est en outre une étape très contrôlée, au niveau du positionnement/choix des ORI, ainsi qu'au niveau de leur allumage coordonné.

D'abord expérimenté au niveau de certains loci spécifiques, le programme d'initiation de la réplication a ensuite été étudié sur l'ensemble du génome humain. Dans la prochaine section, je détaillerai ces démarches ayant permis d'appréhender le déroulement de la réplication sur le génome entier. L'une d'entre-elle (établissement du timing de réplication) sera largement utilisée dans la suite de ce manuscrit pour établir les liens entre organisation nucléaire et déroulement de la réplication.

2.2- Approches génomiques et à haut-débit pour l'étude de la réplication

Il existe plusieurs manières d'étudier la réplication avec des approches génomiques et haut-débit. Nous en détaillerons 4 familles dans les sections suivantes : l'étude de biais de composition en nucléotides dans le génome laissés par la ré-

2.2- Approches génomiques et à haut-débit pour l'étude de la réplication

plication, l'identification des régions où se fixent les complexes pré-répliatif, l'identification des régions fonctionnant réellement comme des ORI, et l'élucidation du programme temporel de réplication (séquences se répliquant en début ou en fin de phase S).

Dans cette section, l'accent sera principalement placé sur les études effectuées chez les métazoaires, lorsqu'elles existent.

2.2.1- Biais de composition en nucléotides et réplication

La complémentarité des deux brins d'ADN veut que le nombre d'adénine $[A]$ et le nombre de thymine $[T]$ soient identiques lorsqu'on considère les deux brins d'ADN (règle de parité #1, Chargaff). Maintenant, si l'on suppose un modèle de mutation neutre, le taux de substitution devrait être globalement identique sur les deux brins : on accumule donc autant de T sur le brin 1 que de T sur le brin 2, soit $[T]_{(1)} = [T]_{(2)}$, en notant (1) et (2) les deux brins d'ADN. Combiné avec la règle de parité #1 ($[T]_{(2)} = [A]_{(1)}$), on obtient donc que, sous réserve de mutations neutres, le nombre de A et de T est globalement identique sur chaque brin d'ADN, soit $[T]_{(1)} = [A]_{(1)}$ (règle de parité #2, [Sueoka, 1995; Lobry, 1995]).

La règle de parité #2 est valable lorsqu'on considère de grandes portions génomiques (par exemple un chromosome entier), mais si l'on se focalise sur de plus petites régions, on peut observer une rupture locale de cette règle. Pour étudier ces ruptures, on définit localement les trois variables suivantes qui permettent d'évaluer ce biais de composition :

$$S_{GC} = \frac{[G] - [C]}{[G] + [C]}, \quad S_{TA} = \frac{[T] - [A]}{[T] + [A]} \quad \text{et} \quad S = S_{GC} + S_{TA}$$

Si localement, le nombre de A et le nombre de T est identique, on devrait avoir $S_{TA} = 0$. Néanmoins, les processus nucléaires peuvent localement biaiser le modèle de mutation neutre et conduire à des asymétries en nucléotides.

1. Lors de la transcription, les deux brins ne sont pas symétriques : le brin matrice formera un hybride ADN-ARN dans la bulle de transcription alors que le brin non-codant restera sous forme simple brin. De plus, les deux brins n'établiront pas des contacts similaires avec la machinerie de transcription. Ces dissymétries peuvent conduire à des patrons de mutations différents sur les deux brins au niveau des gènes exprimés.
2. Lors de la réparation de l'ADN, et notamment pour la réparation couplée à la transcription, les deux brins ne sont là encore pas symétriques, ce qui peut amplifier la dissymétrie déjà mise en place par la bulle de transcription.
3. Enfin, en cours de réplication, les deux brins ne sont pas non plus symétriques, puisque l'un sera répliqué de manière directe et l'autre de manière discontinue avec la formation de fragments d'Okazaki. Cette dissymétrie peut conduire à un biais de mutation avec un gain en G et en T sur le brin direct [Frank and Lobry, 1999].

L'étude de ce biais de composition peut permettre de détecter des origines de réplication. En effet, au niveau des ORI, le biais de composition associé à la réplication doit changer brutalement puisqu'en 5' de l'ORI, la réplication a lieu de manière directe, alors qu'elle s'effectue de manière discontinue en 3'. On accumulera donc plus de G et de T en 5' de l'ORI, et plus de A et de C en 3'. La détection

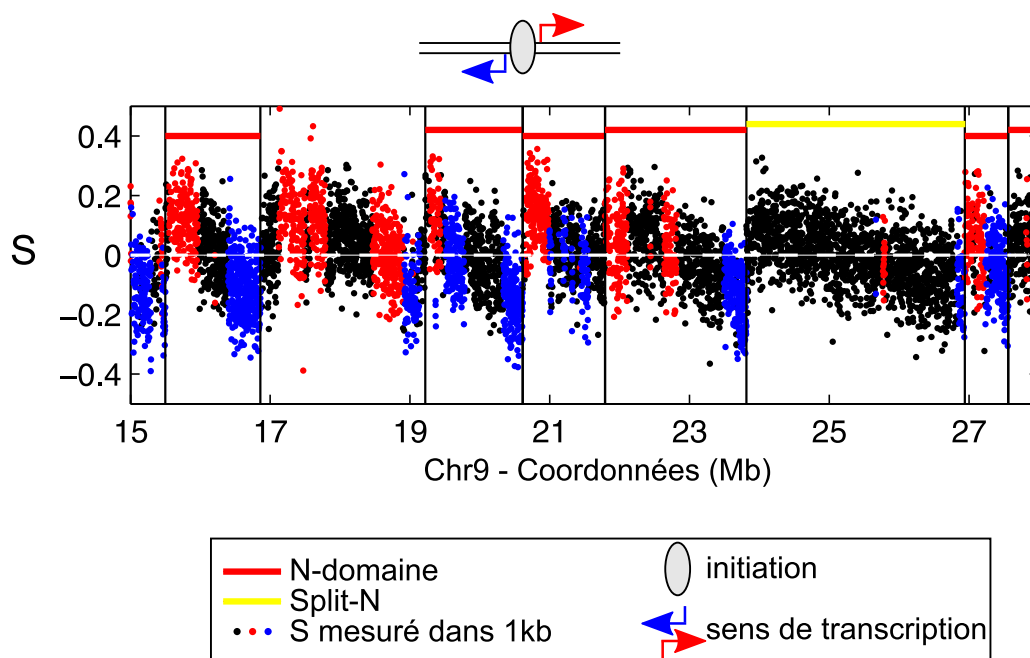


FIGURE 18 – Détection de N-domaines de timing d'après les biais de composition en nucléotides. Le biais de composition en GC et TA (S) est mesuré le long du génome par fenêtre de 1 kb. Chaque point représente une mesure (point noir = inter-génique, point rouge = (+) gène, transcrit vers la droite, point bleu = (-) gène, transcrit vers la gauche). Les barres horizontales rouges correspondent aux N-domaines identifiés, les barres noires verticales symbolisent les bordures des N-domaines. La barre horizontale jaune correspond à un N-domaine dont le cœur ne présente pas une décroissance linéaire. Les bordures des N-domaines correspondent à des zones d'initiation (voir texte) autour desquelles les gènes sont globalement divergents.

de changements brusques de biais de composition permet avec succès de positionner l'ORI de différentes bactéries [Lobry, 1996]. Appliquée chez l'homme, cette méthode a permis de détecter ≈ 1000 changements brusques de biais de composition, permettant non seulement de retrouver la plupart des 10 ORI humaines connues jusqu'alors, mais surtout d'identifier nombre d'ORI potentielles [Brodie of Brodie et al., 2005; Touchon et al., 2005]. Entre ces ORI potentielles, le biais de composition décroît linéairement, ce qui pourrait s'expliquer par l'activation d'ORI intermédiaires à des positions différentes d'une cellule à l'autre. Le biais de composition décrit donc dans son ensemble une forme de N. Approximativement 700 N-domaines d'une taille moyenne de 1,2 Mb ont pu être détectés dans le génome. Au niveau de leurs bordures, qui prédisent le positionnement de zones d'initiation, les gènes sont globalement divergents : les N-domaines pourraient correspondre à des unités fonctionnelles où la réplication et la transcription seraient coordonnées pour éviter les collisions entre les machineries de réplication et de transcription [Huvet et al., 2007] (figure 18). Ces bordures correspondent en outre à des zones de chromatine ouverte (hyper-sensibles à la DNase, hypométhylées sur les CpG, et marquées par la présence de H3K4me3) [Audit et al., 2009].

Cette méthode permet donc de détecter des ORI potentielles grâce à l'empreinte laissée par la réplication sur la séquence. Pour que cette empreinte subsiste dans le génome, il faut qu'elle soit transmise de génération en génération : les ORI identifiées par cette méthode ne peuvent donc correspondre qu'à des zones d'initiation dans la lignée germinale, la seule dont le génome est transmis à la génération future.

2.2- Approches génomiques et à haut-débit pour l'étude de la réplication

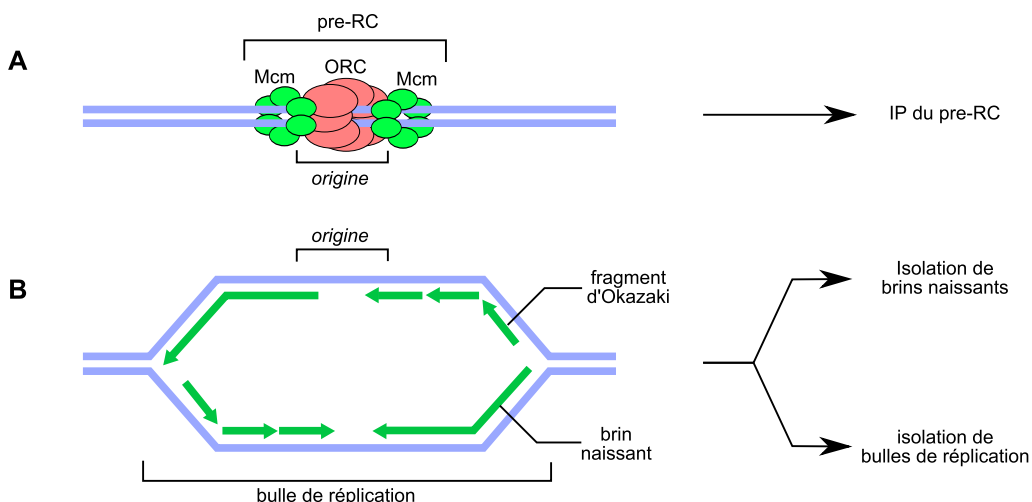


FIGURE 19 – Différentes méthodes permettant la détection d'origines de réplication. A- ChIP dirigé contre le pre-RC. B- Isolation de brins naissants ou de bulles de réplication.

Cette analyse du biais de composition en GC et AT permet de prédire, d'après la séquence, le positionnement d'origines de réplication. Néanmoins, les domaines identifiés sont trop grands pour être répliqués par les deux seules fourches émanant des bordures du N-domaine. D'autres ORI intermédiaires sont réparties à l'intérieur de ces N-domaines. Des méthodes expérimentales ont alors été mises au point pour identifier, sur l'ensemble du génome, les sites où s'initie la réplication.

2.2.2- Identification de sites de liaison du complexe pré-réplicatif (pre-RC)

Le complexe pré-réplicatif (pre-RC) est composé des protéines Orc et des protéines Mcm (voir page 60). L'analyse par immunoprécipitation de chromatine (ChIP) des sites de fixation du pre-RC apporte des informations sur le positionnement des zones où démarre la réplication (figure 19A). Ainsi, environ 360 sites ont été identifiés chez *S. cerevisiae* sur lesquels se lient conjointement 3 protéines Mcm et les protéines Orc [Wyrick et al., 2001]. Ces sites se trouvent majoritairement dans des régions inter-géniques. Ces conclusions ont été complétées par les travaux de Xu et al. [2006b] dans lesquels a été déterminé le motif d'ADN reconnu par le pre-RC. Chez *S. pombe*, Orc1 et Mcm6 sont également plutôt retrouvés entre les régions géniques et 84% des sites de fixation d'Orc1 coïncident avec ceux de Mcm6, permettant l'identification de 460 sites de fixation du pre-RC [Hayashi et al., 2007].

La densité en nucléosome au niveau des sites de fixation des complexes Orc a également été analysée chez *S. cerevisiae*. Si une déplétion en nucléosome est toujours observée au niveau des motifs consensus ARS (du fait de leur forte teneur en AT), cette déplétion est beaucoup plus prononcée lorsque le motif ARS correspond également à un site de fixation des protéines Orc [Eaton et al., 2010].

Chez la drosophile, une stratégie similaire a été mise en place pour identifier les sites de fixation des complexes Orc, d'abord sur le chromosome 2L [MacAlpine et al., 2004], puis sur l'ensemble du génome [MacAlpine et al., 2010; Eaton et al., 2011]. Les deux tiers des sites de fixation des complexes Orc identifiés coïncident

avec le début des gènes (TSS), et principalement avec le TSS des gènes exprimés [MacAlpine et al., 2010]. De plus, ces sites Orc, y compris pour ceux hors des TSS, correspondent à des régions pauvres en nucléosomes, marquées par un enrichissement en H3.3 et où le renouvellement des nucléosomes est globalement rapide [MacAlpine et al., 2010; Eaton et al., 2011]. Cette signature chromatinienne des sites initiateurs de la réplication est tellement particulière qu'elle permet, en étant combinée à d'autres marques épigénétiques, d'établir un modèle prédictif pour identifier ces sites où démarre la réplication [Eaton et al., 2011]. Chez les métazoaires, une ORI serait donc principalement définie par sa structure chromatinienne.

A ma connaissance, aucune détection de régions initiatrices n'a été reportée chez les mammifères grâce à l'analyse par ChIP du pre-RC sur l'ensemble du génome. Les raisons ne sont pas complètement claires mais pourraient à la fois avoir une origine biologique (les protéines Orc ne sont pas exclusivement impliquées dans la réplication et peuvent se fixer à l'ADN en dehors des zones initiatrices de la réplication) et résulter de problèmes techniques (spécificité des anticorps, rapport signal sur bruit) [Schepers and Papior, 2010]. Pour pallier aux problèmes de spécificité/sensibilité, Lubelsky et al. [2011] ont biotinylé *in-vivo* les protéines Orc et Mcm pour analyser leurs distributions par précipitation de la biotine sur de la chromatine cross-linkée. Le pre-RC ayant été retrouvé au niveau des ORI connues du locus DHFR des cellules CHO, l'extension de cette étude pourrait permettre d'identifier sur l'ensemble du génome les sites de fixation du pre-RC. Notons tout de même qu'un ChIP dirigé contre Orc2 et analysé sur l'ensemble du génome de la souris a été effectué par Cayrou et al. [2011]. Les pics obtenus coïncident avec des zones d'initiation, indiquant que, même si le ChIP n'est peut-être pas analysable pour identifier des ORI, un enrichissement en Orc est observé au niveau des zones d'initiation.

Efficace chez la levure et la drosophile, l'immuno-précipitation de sites de fixation du pre-RC ne permet pas aisément de détecter des ORI chez les mammifères. D'autres stratégies ont donc été développées pour directement capturer les événements d'initiation.

2.2.3- Détection de sites d'initiation

L'une des stratégies utilisées pour détecter des événements d'initiation consiste à isoler des brins naissants d'ADN (figure 19B). Les brins naissants sont synthétisés sur le brin précoce au niveau d'une bulle de réplication et commencent par une courte amorce ARN en 5'. Pour s'approcher au plus près de l'ORI, tout en évitant de co-purifier les fragments d'Okazaki (≈ 200 bp) répartis sur l'ensemble du génome, les fragments de 800-1200 bp sont isolés par centrifugation sur gradient. Pour enrichir en fragments d'ADN résultant réellement d'une bulle de réplication (et non d'une fragmentation secondaire du génome lors du protocole expérimental), deux approches existent : (i) un traitement à la λ -exonucléase peut être réalisé pour digérer toutes les molécules d'ADN ne contenant pas en 5' des nucléotides ARN, ou (ii) les cellules peuvent avoir été cultivées avec du BrdU juste avant l'extraction d'ADN et une immunoprécipitation du BrdU est alors réalisée après la centrifugation sur gradient. Grâce à ces deux stratégies, des ORI ont pu être identifiées chez la souris [Sequeira-Mendes et al., 2009; Karnani et al., 2010; Cayrou et al., 2011] et l'homme [Cadoret et al., 2008; Kar-

2.2- Approches génomiques et à haut-débit pour l'étude de la réplication

nani et al., 2010; Valenzuela et al., 2011; Martin et al., 2011]. Elles sont principalement retrouvées dans les régions riches en GC, et notamment au niveau des îlots CpG. Un enrichissement en ORI est également observé à proximité du TSS des gènes. L'espacement entre les ORI, bien que variable d'une région à l'autre, serait en moyenne de 25 kb [Karnani et al., 2010; Cayrou et al., 2011] ou de 100 kb [Cadoret et al., 2008; Sequeira-Mendes et al., 2009].

Une approche complètement différente pour identifier les zones initiatrices de la réplication est basée sur la capture des bulles de réplication (figure 19B) dans des plots d'agarose [Mesner et al., 2006]. Les bulles de réplication ainsi piégées ont été analysées par une hybridation sur une puce à ADN couvrant la région ENCODE du génome [Mesner et al., 2011]. Comme pour l'isolation de brins naissants, cette étude montre que les ORI sont plus souvent associées à des gènes qu'attendu si leur distribution était aléatoire, et que les cellules HeLa et Gm06990 (lymphoblastoïde) utilisent des jeux d'origines bien distincts.

La reproductibilité des isolations de brins naissants et du piégeage des bulles de réplication peut être estimée. Des isolations de brins naissants ont été effectuées par deux groupes différents dans des cellules HeLa (quoique sur deux clones différents, l'un cultivé en suspension et l'autre adhérent). Pourtant seules 14% des ORI identifiées sont communes aux deux études [Cadoret et al., 2008; Karnani et al., 2010]. La méthodologie basée sur le piégeage de bulles de réplication a également été employée dans des cellules HeLa et, là encore, le pourcentage d'ORI communes aux trois études est faible [Mesner et al., 2011]. Les contrôles étant rigoureux dans les trois études, et les propriétés des ORI identifiées dans les 3 études étant similaires, ce faible pourcentage de recouvrement pourrait être expliqué par un sous-échantillonnage des ORI. Chaque étude aurait pu isoler un jeu incomplet de toutes les ORI : si les plus robustes (comme celle à proximité de c-myc) sont retrouvées dans les trois préparations, la diversité des ORI moins robustes pourrait être sous-évaluée dans chaque préparation. Des études ultérieures seront donc vraisemblablement nécessaires pour identifier toutes les origines utilisables dans la cellule.

Les deux méthodologies existantes (isolation de brins naissants ou piégeage de bulles de réplication) permettent d'identifier de nombreuses zones d'initiation. Elles sous-échantillonnent cependant les ORI potentielles.

2.2.4- Établissement du timing de réplication

Une autre manière d'étudier le déroulement de la réplication avec des approches génomiques est, non plus d'identifier les zones d'initiation, mais de déterminer à quel moment de la phase S un locus donné est répliqué (timing de réplication).

a- Les différentes méthodes

Deux familles de méthodes existent pour établir le timing de réplication : celles basées sur une modification du nombre de copies d'un locus, et celles basées sur une incorporation de BrdU. Dans les deux cas, des cellules en phase S sont préalablement isolées (éventuellement à différentes étapes de la phase S).

Variation du nombre de copies Cette méthode se base sur le fait que lorsqu'un locus est répliqué, il est présent en 2 copies dans la cellule. Analyser le nombre de copie de chaque locus permet donc de savoir s'il est répliqué ou non.

Par FACS, des cellules sont isolées en phase G1 et en phase S. Leur ADN est ensuite purifié et marqué dans deux couleurs différentes avant hybridation sur puce à ADN. Si un locus est répliqué en début de phase S, la plupart des cellules isolées en phase S devrait en contenir 2 copies, alors que si ce locus est répliqué en fin de phase S, seule une fraction des cellules en présentera 2 copies. En moyenne, sur une population cellulaire, compter le nombre de copie d'un locus dans une population de cellules en phase S permet donc d'accéder au timing de réplication. Cette méthode a été employée chez l'homme, d'abord avec une résolution de 1 Mb sur l'ensemble du génome [Woodfine et al., 2004], puis avec une résolution de 100 kb sur le chromosome 6 humain [Woodfine et al., 2005].

Enrichissement en loci marqués par du BrdU Dans la méthode précédente, le signal entre locus répliqué et non-répliqué varie au maximum d'un facteur 2. Cet enrichissement d'un facteur 2 est relativement faible et d'autres méthodologies ont été développées pour l'augmenter. Pour cela, du BrdU est incorporé dans l'ADN en cours de synthèse, puis les cellules sont triées par FACS de manière à isoler celles du premier tiers de la phase S et celles du dernier tiers. L'ADN de ces deux populations est extrait et la fraction en cours de réplication, marquée au BrdU, est isolée par immuno-précipitation ou centrifugation sur gradient. Cette fraction, enrichie en séquences en cours de réplication, est ensuite analysée par puce à ADN ou séquençage haut-débit. L'abondance relative de chaque locus du génome dans la fraction "début" ou "fin de phase S" déterminera s'il est répliqué précocement ou tardivement. Cette méthode a été appliquée chez la drosophile [Schübeler et al., 2002; MacAlpine et al., 2004], chez la souris [Hiratani et al., 2008, 2010] et chez des cellules en cultures humaines [White et al., 2004; Ryba et al., 2010].

Afin de gagner en résolution temporelle, la phase S peut être subdivisée, non plus en trois fractions, mais en 4 à 6 fractions pour purifier l'ADN répliqué en tout-début, début, milieu, fin ou toute-fin de phase S. Cette approche à haute résolution temporelle a été utilisée sur des cellules humaines [Hansen et al., 2010; Chen et al., 2010] et murines [Farkash-Amar et al., 2008]. A de nombreuses reprises dans la suite de ce manuscrit, nous ferons référence au timing de réplication établi par Hansen et al. [2010].

Il existe deux méthodologies pour établir le timing de réplication. Nous allons maintenant détailler quelles informations sur le déroulement de la phase S ont été obtenues grâce à ces méthodologies.

b- Propriétés des régions précoces et des régions tardives

Chez la drosophile, la souris et l'homme, les régions répliquées en début de phase S correspondent à des portions de chromatine riches en GC, denses en gènes et exprimées [Schübeler et al., 2002; MacAlpine et al., 2004; Woodfine et al., 2004; White et al., 2004; Hiratani et al., 2008; Farkash-Amar et al., 2008; Ryba et al., 2010]. A l'inverse, les zones tardives sont principalement retrouvées au niveau des régions pauvres en gènes, peu exprimées, et à teneur en GC plus basse. Concernant les modifications d'histones, les régions marquées par une forte teneur en H3K4me1-3,

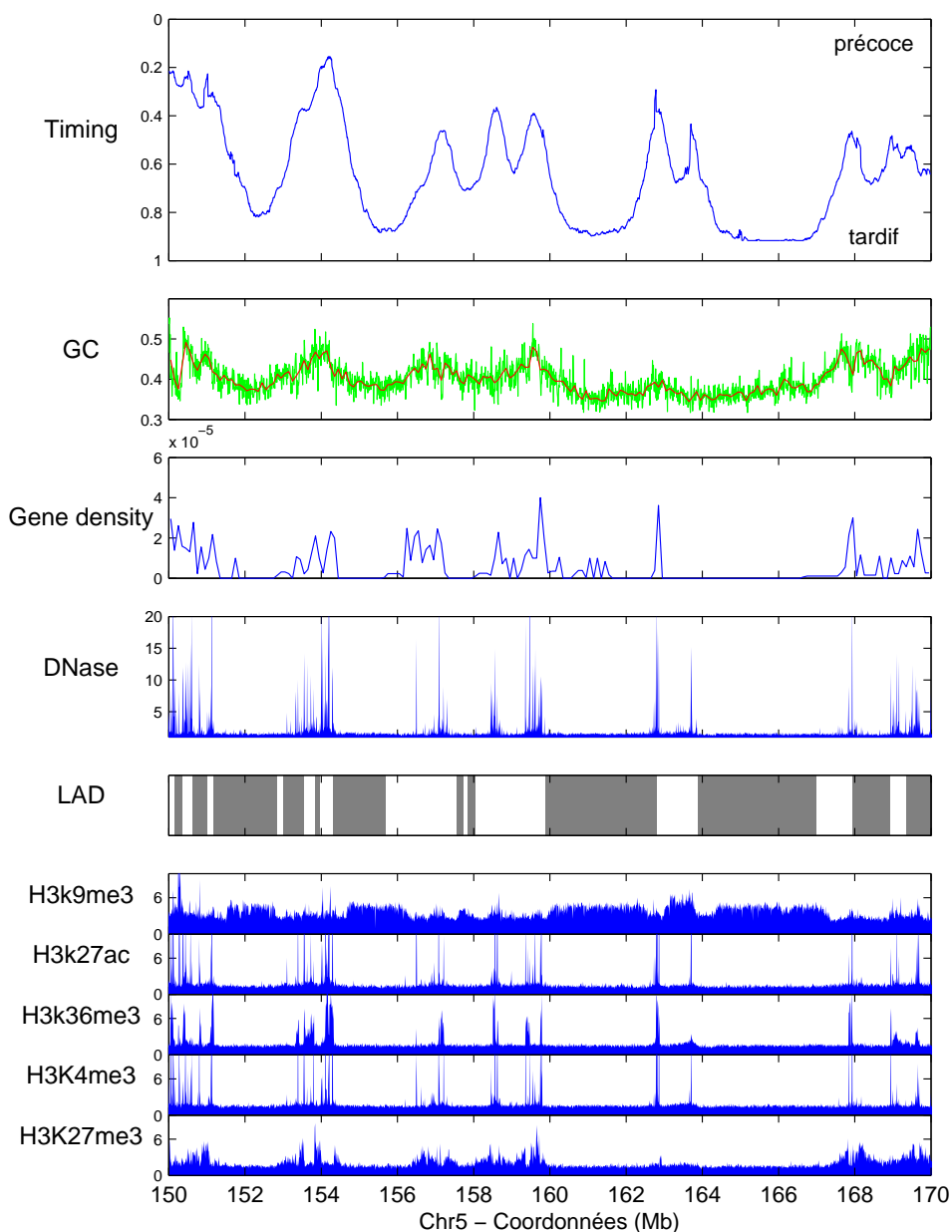


FIGURE 20 – Propriétés des régions répliquées en début de phase S (précoces) et des régions répliquées en fin de phase S (tardives) dans les cellules humaines K562. Sur cette figure, sont représentés de haut en bas le long de 20 Mb du chromosome 5 humain : le timing de réplication (précoce en haut du panneau, tardif en bas) obtenu par [Hansen et al. \[2010\]](#) et post-traité par [Baker et al. \[2012\]](#) (voir aussi le paragraphe U-domaine ci-après) ; la teneur en GC (sur 10 kb en vert et sur 100 kb en rouge) ; la densité en gènes par paire de base (calculée à partir de RefSeq de UCSC sur une fenêtre de 100 kb) ; la couverture en sites hypersensibles à la DNase I (moyenne sur 10 kb à partir des données UCSC UW DNaseI HS) ; la présence de domaines d'interaction avec la lamina nucléaire (LAD [[Guelen et al., 2008](#)]) ; et enfin 5 modifications d'histones (moyenne sur 10 kb à partir des données UCSC Broad Histones). Le timing de réplication, les profils de marques d'histones et de sites hypersensibles à la DNase ont été établis dans les cellules K562 (issue d'une leucémie myéloïde à propriétés érythrocytaire), les LAD proviennent de fibroblastes. Les régions précoces sont riches en GC, denses en gènes, marquées par la présence de sites hypersensibles à la DNase, et riches en H3K4me3, H3K27Ac, H3K36me3. A l'inverse les régions tardives sont appauvries en gènes, à basse teneur en GC, riches en H3K9me3, et présentent des interactions avec les lamines nucléaires. Sur cette portion du génome, H3K27me3 est plutôt retrouvée au niveau des régions précoces.

H3K9ac, H3K27ac, H3K36me3 sont plutôt répliquées en début de phase S, tandis que celles marquées par H3K9me2,3, au moins chez l'homme, sont plutôt répliquées en fin de phase S [Hiratani et al., 2008; Ryba et al., 2010]. En somme, ces études à haut-débit confirment que les régions ouvertes et exprimées sont répliquées en début de phase S alors que la chromatine répressive non exprimée est répliquée en fin de phase S (figure 20). Notons néanmoins que la marque répressive H3K27me3, empreinte d'une répression par les complexes Polycomb, peut être retrouvée aussi bien au niveau des régions actives que répressives [Ryba et al., 2010].

Chez la drosophile, ces relations entre une réplication précoce et un taux de transcription ou une densité en gènes élevée sont plus fortes lorsqu'elles sont intégrées sur 100 kb (plusieurs gènes) que lorsqu'elles sont mesurées sur 10 kb (un seul gène) [MacAlpine et al., 2004]. Ceci suggère que le lien transcription \leftrightarrow réplication reflète la contribution de plusieurs gènes. Chez la souris, cette corrélation est également plus forte lorsqu'elle est intégrée sur de larges fenêtres (600 kb) [Hiratani et al., 2008]. Cette taille de fenêtre est du même ordre de grandeur que celle des domaines chromatinien structuraux et fonctionnels (voir page 44), indiquant que le timing de réplication reflète un état moyenné de la chromatine à l'échelle d'un domaine.

Cette relation entre transcription et réplication n'est pas observée chez la levure [Raghuraman et al., 2001], indiquant qu'il pourrait s'agir d'une spécificité des métazoaires. De plus, ces tendances générales ne doivent pas effacer la diversité des cas particuliers. Par exemple, au moins chez la drosophile, des régions précoces sont observées dans la région hétérochromatique à proximité du centromère du chromosome 2L [Schübeler et al., 2002] et dans des zones relativement appauvries en sites de fixation de l'ARN-polymérase II [MacAlpine et al., 2004]. Enfin, la valeur du timing de réplication est une valeur moyenne : grâce à leur expérience à haute résolution temporelle, Farkash-Amar et al. [2008] et Hansen et al. [2010] ont pu identifier des loci présentant deux vagues d'initiation qui s'expliqueraient, au moins en partie, par à des initiations asynchrones sur les deux allèles.

Les loci répliqués en début de phase S et en fin de phase S sont majoritairement retrouvés dans des contextes chromatinien différents. L'euchromatine est globalement précoce, alors que l'hétérochromatine est répliquée en fin de phase S.

c- Conservation et modification du timing de réplication

L'étude du timing de réplication dans plusieurs types cellulaires a permis d'évaluer son niveau de conservation. En comparant 3 lignées de cellules ES murines, Hiratani et al. [2008] ont remarqué que le timing est extrêmement bien conservé au sein d'un même type cellulaire. Un très bon niveau de conservation est en outre observé entre les régions homologues de l'homme et de la souris pour des cellules ayant la même origine (lymphoblastes [Farkash-Amar et al., 2008; Yaffe et al., 2010], cellules souches [Ryba et al., 2010] et fibroblastes [Yaffe et al., 2010]). Cette conservation n'est pas simplement due à la conservation, entre régions homologues, du taux en GC. Pour aboutir à cette conclusion, Yaffe et al. [2010] ont calculé, pour chaque 50 kb du génome, un "timing de réplication résiduel". Celui-ci est obtenu en soustrayant au timing mesuré sur ces 50 kb, la valeur moyenne

2.2- Approches génomiques et à haut-débit pour l'étude de la réplication

du timing de réplication de toutes les régions du génome ayant la même teneur en GC que les 50 kb analysés. En somme, ces auteurs ont soustrait au timing de réplication, la contribution moyenne du GC. Or le timing de réplication résiduel est toujours corrélé entre les régions homologues humaines et murines, démontrant que la conservation du timing de réplication ne s'explique pas uniquement par la conservation de la teneur en GC.

Par contre, le niveau de conservation est moindre lorsqu'on considère des cellules ayant des origines différentes [Desprat et al., 2009; Hiratani et al., 2010; Hansen et al., 2010; Yaffe et al., 2010; Ryba et al., 2010; Chen et al., 2011]. Hansen et al. [2010] rapportent en effet que 49% du génome montre un timing variable quand celui-ci est comparé entre 4 types cellulaires différents. De même, la comparaison de cellules souches et de cellules différenciées montre qu'entre 20% [Desprat et al., 2009] et 45% [Hiratani et al., 2008] du génome change de timing de réplication lors de la différenciation. Ces changements de timing s'établissent progressivement au cours de la différenciation et sont le plus souvent associés à des changements de niveau d'expression [Hiratani et al., 2010]. Les loci montrant un timing plus tardif dans les cellules différenciées (EtoL, early-to-late) sont majoritairement réprimés, tandis que ceux montrant un timing plus précoce dans les cellules différenciées (LtoE, late-to-early) sont au contraire majoritairement activés.

Certains loci ont un timing de réplication stable d'un type cellulaire à l'autre, tandis que d'autres possèdent un timing qui varie selon le type cellulaire. Des changements de timing s'effectuent au cours de la différenciation, en lien étroit avec les changements d'expression génique. En plus d'avoir été comparé d'une lignée à l'autre, le signal "timing de réplication" a également été segmenté de telle sorte à identifier des domaines fonctionnels.

d- Identification de domaines fonctionnels de timing

Tout comme les profils de modifications d'histones ou d'expression avaient permis d'identifier des domaines fonctionnels, l'étude du timing de réplication a également permis la définition de domaines fonctionnels de réplifications. Deux segmentations ont été proposées : l'une basée sur une succession de domaines à timing constant, et l'autre basée sur des domaines à timing en forme de U. Nous allons détailler ces deux approches dans les paragraphes suivants.

Segmentation en domaines possédant un timing de réplication homogène sur quelques centaines de kb (segmentation à timing homogène) Une segmentation du timing de réplication a été effectuée de manière à identifier des portions de génome dans lesquelles le timing de réplication est homogène sur quelques centaines de kb. La taille de telles régions est comprise entre 200 kb et quelques Mb chez la souris, avec une moyenne à 1,5 Mb et une médiane aux alentours de 1 Mb. Elles peuvent être répliquées tout au cours de la phase S, avec une préférence pour le tout début de la phase S ou la toute fin de la phase S pour les plus grandes [Hiratani et al., 2008]. Une étude indépendante confirme que 90% du génome de la souris correspond à des régions à timing de réplication constant [Farkash-Amar et al., 2008]. Des conclusions similaires ont été obtenues chez l'homme [Ryba et al., 2010]. Les régions de génome à timing homogène s'enchaînent dans le génome et sont reliées par des segments linéaires d'une taille de 0,1 à 0,6 Mb appelées "Régions de transition" [Hiratani et al., 2008].

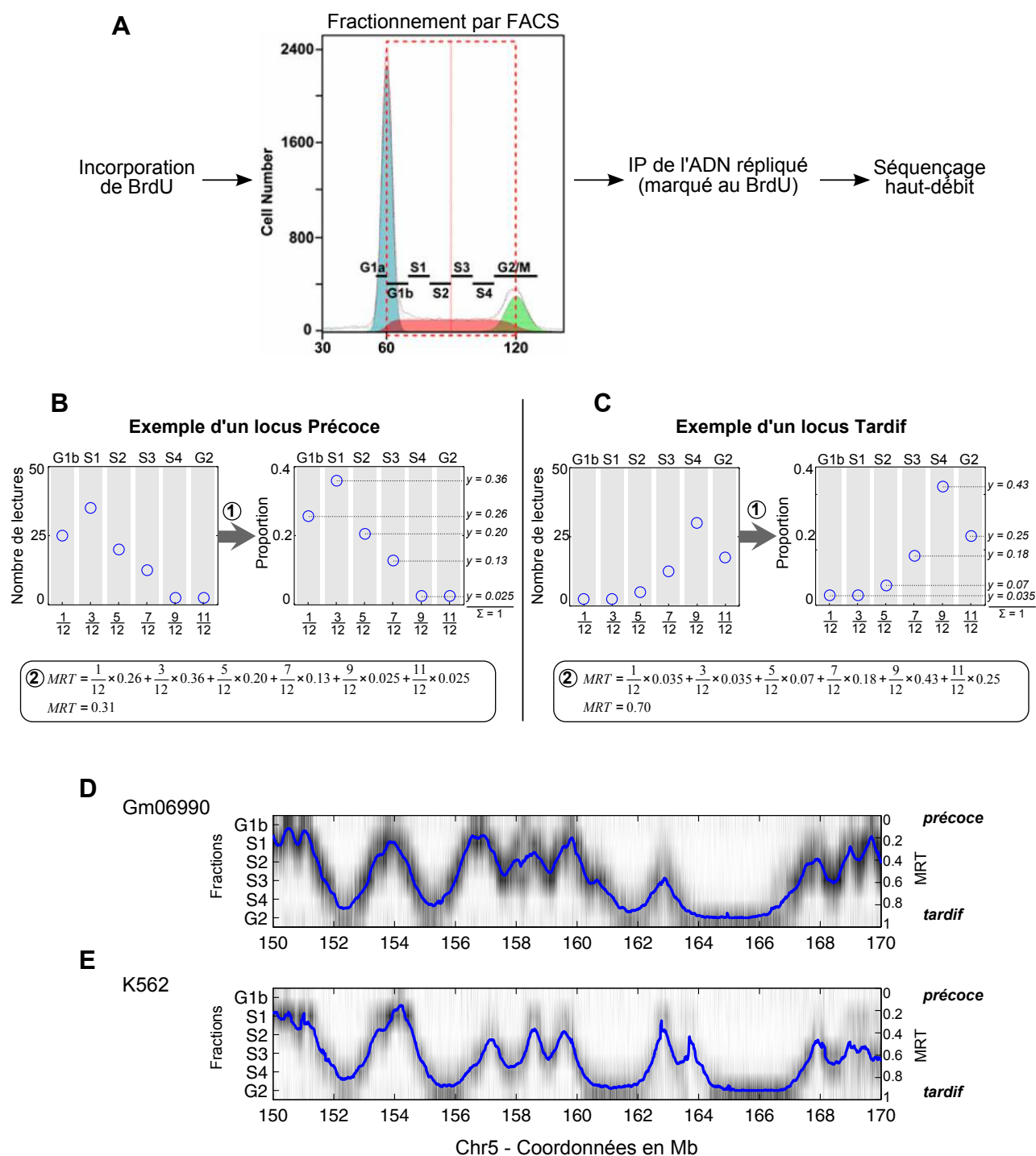


FIGURE 21 – Calcul du timing moyen de réplication à partir des données de Repli-Seq [Hansen et al., 2010]. **A**- Fractionnement de la phase S en 4 fractions (S1, S2, S3 et S4) et isolation de la phase G1b et G2 d'après le profil de FACS. Au final, les fractions retenues sont localisées dans le rectangle pointillé rouge. Une IP anti-BrdU est ensuite réalisée sur l'ADN extrait des cellules de chaque fraction, puis l'ADN correspondant est séquençé. **B**- **C**- Pour chaque locus du génome, on mesure le nombre de lectures sur une fenêtre de 100 kb (gauche). A partir de cette valeur, on calcule la proportion de lectures dans chaque fraction (G1b...G2), de telle sorte que la somme dans toutes les fractions vaille 1 (droite). On quantifie ensuite le timing de réplication moyen (MRT) en calculant la moyenne du temps associé à chaque fraction ($\frac{1}{12}$ pour G1b, $\frac{3}{12}$ pour S1, $\frac{5}{12}$ pour S2, ..., $\frac{11}{12}$ pour G2) pondérée par la proportion de lectures de chaque fraction. Le MRT d'un locus précoce (**B**) sera plus petit que celui d'un locus tardif (**C**). **D**- Profile de timing (MRT, courbe bleue) pour les cellules Gm06990 sur 20 Mb du chromosome 5 humain et comparaison avec les données brutes de Hansen et al. [2010] (pour chaque fraction, densité en lectures par 10 kb, représentée avec une table de correspondance blanc-noir). **E**- Comme D mais pour les cellules K562.

2.2- Approches génomiques et à haut-débit pour l'étude de la réplication

Avec la séparation des cellules en phase S en plusieurs fractions, [Hansen et al. \[2010\]](#) ainsi que [Chen et al. \[2010\]](#) observent que certaines régions ont en effet un timing constant, mais que pour d'autres, le timing suit un continuum de valeurs différentes (timing en forme de U). Ces régions en U étant présentes dans de nombreux types cellulaires, une autre segmentation a été proposée afin de détecter ces U-domaines.

Segmentation en U-domaines La segmentation proposée par [Baker et al. \[2012\]](#) comporte deux étapes.

1. Définition d'un timing de réplication moyen (MRT, Mean Replication Timing) à partir des données de Repli-Seq [[Chen et al., 2010](#); [Hansen et al., 2010](#); [Baker et al., 2012](#)].

A partir du nombre de lectures par 100 kb obtenu des données brutes de séquençage haut-débit, un filtrage est effectué pour en réduire le bruit. Pour chaque locus, on quantifie ensuite la proportion de lectures dans chaque fraction : la somme sur l'ensemble des fractions vaut donc 1. Puis, le MRT du locus est estimé en calculant la moyenne du temps associé à chaque fraction pondérée par la proportion de lectures (figure 21A-C). Cette valeur, qui varie par construction entre 0 (locus répliqué en début de phase S) et 1 (locus répliqué en fin de phase S), est calculée tout le long du génome (figure 21D-E). Par convention, tout au long de ce manuscrit, le timing moyen de réplication (MRT) sera représenté avec un axe des ordonnées dirigé vers le bas (0 = précocité, 1 = tardif, voir figure 21D-E). Pour plus de détail sur le calcul du MRT, voir [Baker et al. \[2012\]](#).

2. D'après le signal MRT, certaines régions précoces peuvent être relativement fines (figure 21D, 154 Mb et 157 Mb par exemple), et une segmentation à timing constant au niveau de ces régions serait peu pertinente. La segmentation développée par [Baker et al. \[2012\]](#) vise donc à détecter les domaines en forme de U comme celui entre les positions 154 et 157 Mb (figure 21D). Ces domaines possèdent donc un cœur répliqué tardivement et sont entourés par deux zones d'initiation précocité. La présence de domaines en forme de U a été détectée sur l'ensemble du génome [[Baker et al., 2012](#)] : on en trouve 882 dans les cellules Gm06990, couvrant 50% du génome avec une taille moyenne de 1,52 Mb. Dans les cellules K562, les 876 U-domaines détectés couvrent 46% du génome et ont une taille moyenne de 1,42 Mb. Les 50% restant du génome correspondent à de grandes régions à timing constant répliquées en début de phase S ou en fin de phase S.

Ces deux segmentations traduiraient deux coordinations différentes de l'allumage des ORI

Dans la version "à timing homogène", la segmentation identifie des domaines où la réplication s'initie peu ou prou en même temps. Comme ces domaines font ≈ 1 Mb et que les ORI sont en moyenne distantes de 25-100 kb, la réplication à timing constant de ces domaines suggère un allumage synchrone de 10 à 40 ORI. A l'inverse, la présence de U-domaines (≈ 1 Mb) indique plutôt le démarrage non synchrone de 10 à 40 ORI, les uns après les autres. Ces deux types de domaines pourraient donc correspondre à des unités fonctionnelles au sein desquelles le programme spatio-temporel de la réplication serait différemment coordonné.

Ces deux modes d'allumage des ORI pourraient coexister dans le génome puisque les U-domaines détectés ne couvrent que 50% du génome, et que les 50% restant correspondent à de grandes régions répliquées soit en début, soit en

fin de phase S. Il est donc possible que la moitié du génome soit répliquée avec une activation des origines les unes après les autres, tandis que la moitié restante soit composée de domaines de timing dans lesquels toutes les origines s'allument de manière synchrone au début ou à la fin de la phase S.

Des domaines de timing ont été identifiés à partir du timing de réplication. Selon la forme du domaine (plat ou en forme de U), la coordination de l'allumage des origines de réplication semble différente. Dans la partie résultat de ce manuscrit, les liens entre organisation à grande échelle de la chromatine et timing de réplication seront analysés. La région du génome majoritairement étudiée dans la partie résultat présente un timing de réplication constitué de domaines en U successifs (chr5:150-170Mb).

Conclusion

Les approches à haut-débit ont permis de préciser sur, l'ensemble du génome, le positionnement des zones d'initiation et d'appréhender le déroulement du programme temporel de la réplication. Les séquences se répliquant en début ou en fin de phase S sont ainsi connues avec une précision de l'ordre de 100 kb. Le timing de réplication est d'autant plus conservé que les types cellulaires sont proches, même entre différentes espèces. L'analyse attentive du timing de réplication a permis de caractériser les modifications d'histones associées aux régions répliquées en début ou fin de phase S, et d'identifier des domaines fonctionnels.

Comme évoqué dans le chapitre précédent, le noyau est un organite hautement organisé, avec la formation de repliements contrôlés et de domaines structuraux. Si nous avons insisté, dans le chapitre précédent, sur les liens entre organisation de la chromatine et transcription, nous allons montrer dans la section suivante que des liens étroits existent également entre organisation de la chromatine et réplication.

2.3- Structuration de la chromatine associée à la réplication

Tout comme la transcription, la réplication est associée à une organisation nucléaire caractéristique. La matrice nucléaire et les boucles de chromatine, qui sont impliquées dans la transcription, jouent également un rôle dans la réplication. Nous verrons tout d'abord que la réplication de l'ADN est associée à la formation de boucles de chromatine de l'ordre de quelques Mb, puis qu'elle se déroule au sein de foyer de réplication. Dans une dernière partie, nous décrirons les résultats d'études récentes qui tendent à montrer, par des méthodologie de la famille "3C", que des contacts semblent s'établir entre régions répliquées de manière synchrone. Nous insisterons alors sur les limites des conclusions tirées et sur les questions subsistantes.

2.3.1- Réplication et boucles d'ADN

Préparer des halos nucléaires permet d'étudier, dans son ensemble, les boucles de chromatine présentes dans un noyau. En effet, lors du traitement hypertonique, les protéines du noyau, y compris les histones, sont en grande partie perdues, ce qui permet de débobiner la chromatine. Les séquences d'ADN ancrées au squelette nucléaire (Scaffold / Matrix Attachment Regions, S/MAR) sont alors maintenues au centre du halo, tandis que les séquences non-ancrées sont refoulées en périphérie (voir page 31 et suivantes). La nature des séquences ancrées à la matrice permet donc d'évaluer dans son ensemble les liens éventuels entre organisation nucléaire et réplication.

a- Halo nucléaire et réplication

Avec la préparation de halos nucléaires, la chromatine se sépare en deux : la fraction ancrée à la matrice nucléaire est retrouvée au cœur du halo (au niveau du vestige nucléaire), tandis que la fraction non-ancrée est retrouvée dans les boucles périphériques. L'incorporation de nucléotides radiomarqués suivi de la préparation de halos nucléaires permet alors d'analyser dans laquelle de ces deux fractions la réplication s'effectue. Après un marquage de 1 min à la thymidine tritiée, 80% des sites d'incorporation de radioactivité sont observés au niveau de la matrice nucléaire, alors que seuls 20% sont retrouvés dans les boucles. Une chasse d'au moins 20 min est nécessaire pour augmenter la proportions de sites radioactifs au niveau des boucles du halo [Vogelstein et al., 1980]. Des résultats similaires ont été obtenus avec la préparation de halos nucléaires après incorporation de BrdU [Gerdes et al., 1994]. Ces deux études indiquent donc que la réplication s'effectue au cœur du halo nucléaire, proche de la matrice nucléaire et des points d'ancrage.

Si la réplication s'effectue au niveau des points d'ancrage des boucles de chromatine, on pourrait s'attendre à ce que la taille de ces boucles soit à peu près similaire à la distance entre origines de réplication. En mesurant le diamètre du halo nucléaire chez plusieurs espèces dont la distance inter-ORI est connue grâce à des marquages sur fibre d'ADN étirée, Buongiorno-Nardelli et al. [1982] ont pu observer une étroite corrélation entre la taille maximale du halo (correspondant à l'étirement maximal des molécules d'ADN) et la distance inter-ORI. Un résultat similaire est obtenu dans des conditions de réplication *in-vitro* avec un extrait d'œufs de Xénopus [Lemaitre et al., 2005]. Ces observations confirment donc le lien entre organisation du noyau (ancrage à la matrice nucléaire et formation de boucles d'ADN) et la réplication.

Une étude beaucoup plus récente corrobore cette interprétation [Courbet et al., 2008]. Une mutante de cellule CHO a été sélectionnée pour sa vitesse lente de progression des fourches ($\approx 0,6$ kb/min). Cependant, grâce à une densité en ORI activées plus élevée, la durée de la phase S de ce mutant est similaire à celle de cellules répliquant leur ADN avec des fourches progressant plus rapidement. Ce mutant, lorsqu'il est cultivé dans un milieu supplémenté en adénine et uridine, retrouve une vitesse normale de progression des fourches ainsi qu'une densité en ORI activées plus basse. Des halos nucléaires ont été préparés chez ce mutant cultivé dans les deux conditions. En présence d'adénine et d'uridine (vitesse rapide et densité en ORI basse), la taille du halo est deux fois plus petite que lorsque le mutant est cultivé sans adénine et uridine (vitesse lente et densité en ORI élevée), confirmant le lien étroit entre distance inter-ORI et taille du halo.

nucléaire [Courbet et al., 2008]. Par FISH, cette étude montre en outre que des ORI qui ne sont utilisées que dans le cas où la vitesse de progression des fourches est lente, se localisent au niveau de la matrice nucléaire lorsque les fourches sont lentes, mais se retrouvent en périphérie du halo lorsque les fourches sont rapides. Ce résultat est en accord avec une réplication s'effectuant au niveau de la matrice nucléaire. Plus surprenant, une phase S est nécessaire pour changer la taille du halo en fonction de la présence ou non d'adénine et d'uridine, indiquant que les points d'ancrage de l'ADN dépendent du déroulement de la réplication du cycle cellulaire précédent. Ce résultat atteste donc de l'importance de la réplication dans l'ancrage à la matrice nucléaire et donc dans l'organisation du noyau.

Les boucles d'ADN visualisées par la préparation de halos nucléaires sont intimement liées au déroulement de la phase S : la réplication s'initie près des points d'ancrage à la matrice nucléaire et permet la formation de boucles en accord avec la distance inter-ORI. Un changement dans le déroulement de la phase S (comme une modification de la vitesse de progression des fourches) affectera l'organisation nucléaire du cycle suivant.

b- Matrice et réplication – Correspondance structurale

La matrice nucléaire, les séquences ancrées à la matrice (S/MAR) et la réplication partagent des acteurs en commun.

S/MAR et ORI coïncident Comme les S/MAR (voir page 31), de nombreuses origines de réplication sont riches en AT et correspondent à des régions d'ADN courbes accumulant les contraintes topologiques qui favorisent l'ouverture de la double hélice d'ADN [Anachkova et al., 2005; Aladjem et al., 2006]. Ces contraintes topologiques, notamment les super-tours négatifs, facilitent d'ailleurs le recrutement des protéines Orc chez la drosophile [Remus et al., 2004].

Partageant ces propriétés communes, S/MAR et ORI sont souvent rencontrées aux mêmes endroits. En effet, des régions d'ancrages à la matrice sont retrouvées au niveau de l'ORI à proximité de c-myc (chez la souris et le Xénope [Girard-Reydet et al., 2004]), au niveau de l'ORI à proximité du gène β -globine (chez l'homme [Djeliova et al., 2001b]), au niveau de l'ORI à proximité du gène lamine B2 (chez l'homme, [Lagarkova et al., 1998]), ou bien encore au niveau de l'ORI à proximité du gène HPRT (chez l'homme, [Koina and Piper, 2005]).

Ces quelques exemples (non-exhaustifs) montrent que S/MAR et ORI peuvent partager les mêmes séquences, renforçant l'idée d'un ancrage des ORI au niveau de la matrice nucléaire.

Matrice nucléaire et protéines de réplication La machinerie de réplication et la matrice nucléaire peuvent partager les mêmes acteurs. En effet, la topoisomérase II, qui est entre autre impliquée dans la résolution des contraintes topologiques créées par la réplication, est une des protéines structurales de la matrice nucléaire [Razin, 2001]. En phase M, la topoisomérase II serait d'ailleurs directement impliquée dans le remodelage de la taille des boucles de chromatine et de l'espacement inter-ORI [Lemaitre et al., 2005]. De plus, PCNA, au moins une sous-

2.3- Structuration de la chromatine associée à la réplication

unité des protéines RPA, l'ADN polymérase α ainsi que la primase sont également associées à la matrice nucléaire¹.

Enfin, la protéine Orc1 est majoritairement retrouvée dans la fraction insoluble du noyau résistante à la DNase. En effet, après une extraction au Triton X-100 de noyaux de cellules HeLa (ce qui sépare le noyau en une fraction soluble – non chromatinienne – et une fraction insoluble – chromatine –), suivi d'une digestion à la DNase de la fraction insoluble (ce qui extrait la majeure partie de l'ADN), la protéine Orc1 ségrège dans la fraction insoluble résistante à la DNase [Ohta et al., 2003]. Bien que le protocole expérimental de cette étude ne corresponde pas à celui d'une extraction de matrice nucléaire, la localisation de Orc1 au niveau d'un résidu nucléaire insoluble et résistant à la DNase est en accord avec l'ancrage des acteurs de la réplication sur un squelette nucléaire.

Dans l'ensemble, ces observations montrent d'une part que des régions d'ancrage à la matrice nucléaire sont observées au niveau des origines de réplication, et d'autre part que plusieurs acteurs moléculaires de la réplication sont associés au squelette nucléaire.

c- Matrice et réplication – Correspondances fonctionnelles

Un S/MAR permet le maintien d'un plasmide sous forme épisomale dans des cellules de mammifères En plus de partager des acteurs communs, S/MAR et ORI peuvent avoir des fonctions redondantes. En effet, la présence de S/MAR est essentielle au plasmide pEPI-1 (~6,7 kb) pour assurer son maintien sous forme épisomale et sa transmission sur plus de 30 générations dans des cellules CHO ou HeLa. Sans ce S/MAR, le plasmide est intégré dans le génome des clones sélectionnés pour leur résistance au G418 [Schaarschmidt et al., 2004]. Des résultats similaires ont été obtenus avec la même construction (excepté le remplacement de la GFP par eGFP) dans des cellules primaires [Papapetrou et al., 2006].

Les fourches de réplication sont associées à la matrice nucléaire Avec la préparation de halos nucléaires, nous avons déjà évoqué que la synthèse d'ADN semblait s'effectuer au centre du halo, et non au niveau des boucles de chromatine [Vogelstein et al., 1980]. Ce résultat a été confirmé en ajoutant, après le traitement hypersalin, une digestion à la DNase I [Pardoll et al., 1980]. Ainsi, après un marquage pendant 30 secondes de l'ADN naissant à la thymidine tritiée, il a été montré que 90% de la radioactivité totale est retrouvée sur le vestige nucléaire. Pourtant, seul 15% de l'ADN total, correspondant à celui ancré à la matrice nucléaire, avait résisté au traitement à la DNase. Le traitement hyper-salin (2M) pouvant conduire à des artefacts expérimentaux (notamment à la formation d'agrégats protéiques), Vaughn et al. [1990] ont confirmé ce résultat à l'aide d'extraction nucléaire avec un tampon isotonique (100 mM lithium acatate, 10 mM LIS).

ORI et matrices s'associent lors de la réplication Puisque l'ADN naissant est associé à la matrice nucléaire, les ORI doivent également en être proche lors de la phase S. Pour le reste du cycle, deux possibilités sont envisageables : soit les ORI sont en permanence associées à la matrice nucléaire, soit elles s'y associent juste au moment de la réplication. L'analyse de la distribution de l'ORI du locus DHFR

1. D'après nuclear matrix associated proteins database, NMP-DB, <http://roslab.org/services/nmpdb/index.html>

dans des cellules CHO (toutes phases du cycle confondues) sur des préparations d'halos nucléaires montre que cette ORI se localise à la fois dans la fraction liée à la matrice et dans les boucles de chromatine [Djeliova et al., 2001a]. Ceci démontre donc que les ORI ne sont pas en permanence associées à la matrice nucléaire. Par contre, cette ORI, qui est répliquée en tout début de phase S, montre un ancrage plus fort à la matrice nucléaire en fin de phase G1, et un ancrage plus faible au cours de la phase S ou en début de phase G1 [Djeliova et al., 2001b]. Ce résultat supporte donc une association transitoire des ORI avec le squelette nucléaire, avant leur activation.

La matrice nucléaire supporte l'initiation Si les ORI s'activent en début de phase S s'ancrant à la matrice nucléaire en fin de phase G1, il devrait être possible d'initier la réplication sur des extractions de matrices nucléaires. Pour tester cette idée, des extractions de matrices préparées sur des cellules en fin de phase G1 sont incubées avec un extrait protéique de cellules HeLa en phase S. Avec ce système, Radichev et al. [2005] notent une incorporation de nucléotides radiomarqués, démontrant que les séquences d'ADN ancrées sur la matrice nucléaire se répliquent. Inversement, des préparations de matrice nucléaire effectuées en début de phase G1 sont incapables d'initier la réplication. Ce résultat, même s'il ne démontre pas de lien de causalité, est en accord avec le recrutement des ORI à la matrice nucléaire en fin de phase G1.

De nombreuses études montrent que la réplication s'effectue au niveau de ce squelette résiduel mal connu que l'on obtient après extraction d'une grande partie des protéines nucléaires avec un traitement hypersalin. Les ORI seraient recrutées avant leur activation sur cette matrice nucléaire qui servirait ensuite de support à la réplication (pour de plus amples détails, voir les revues de Anachkova et al. [2005] et de Ottaviani et al. [2008]). En plus de s'associer à la matrice nucléaire, la réplication se déroule avec une organisation spatiale particulière au sein du noyau : elle a lieu au sein de foyers.

2.3.2- Les foyers de réplication

a- La réplication au niveau de foyers discrets

L'utilisation de nucléotides modifiés permet de détecter, dans le noyau, les sites en cours de réplication. Ainsi l'incorporation de biotine-dUTP dans des cellules perméabilisées montre que la réplication s'effectue dans des foyers discrets [Nakayasu and Berezney, 1989]. Des résultats similaires ont été obtenus avec des nucléotides halogénés (avec un groupement fluor FdU, chlore CldU, brome BrdU ou iode IdU) [Nakayasu and Berezney, 1989; O'Keefe et al., 1992] ou portant un groupement ethynyl (EdU) pouvant réagir avec un Alexa Fluor azide [Cseresnyes et al., 2009]. Les foyers de réplication, bien que majoritairement étudiés chez les métazoaires, ont également été observés chez les plantes [Sparvoli et al., 1994], chez *S. cerevisiae* [Pasero et al., 1997] et chez *S. pombe* [Meister et al., 2007].

La distribution ainsi que la taille/forme des foyers changent au cours de la phase S. En effet, en tout début de la phase S, on observe ~1000 petits foyers [Ma et al., 1998] répartis dans l'ensemble du noyau, à l'exclusion des nucléoles et des bords du noyau. Plus on avance dans la phase S, plus les foyers de réplication se

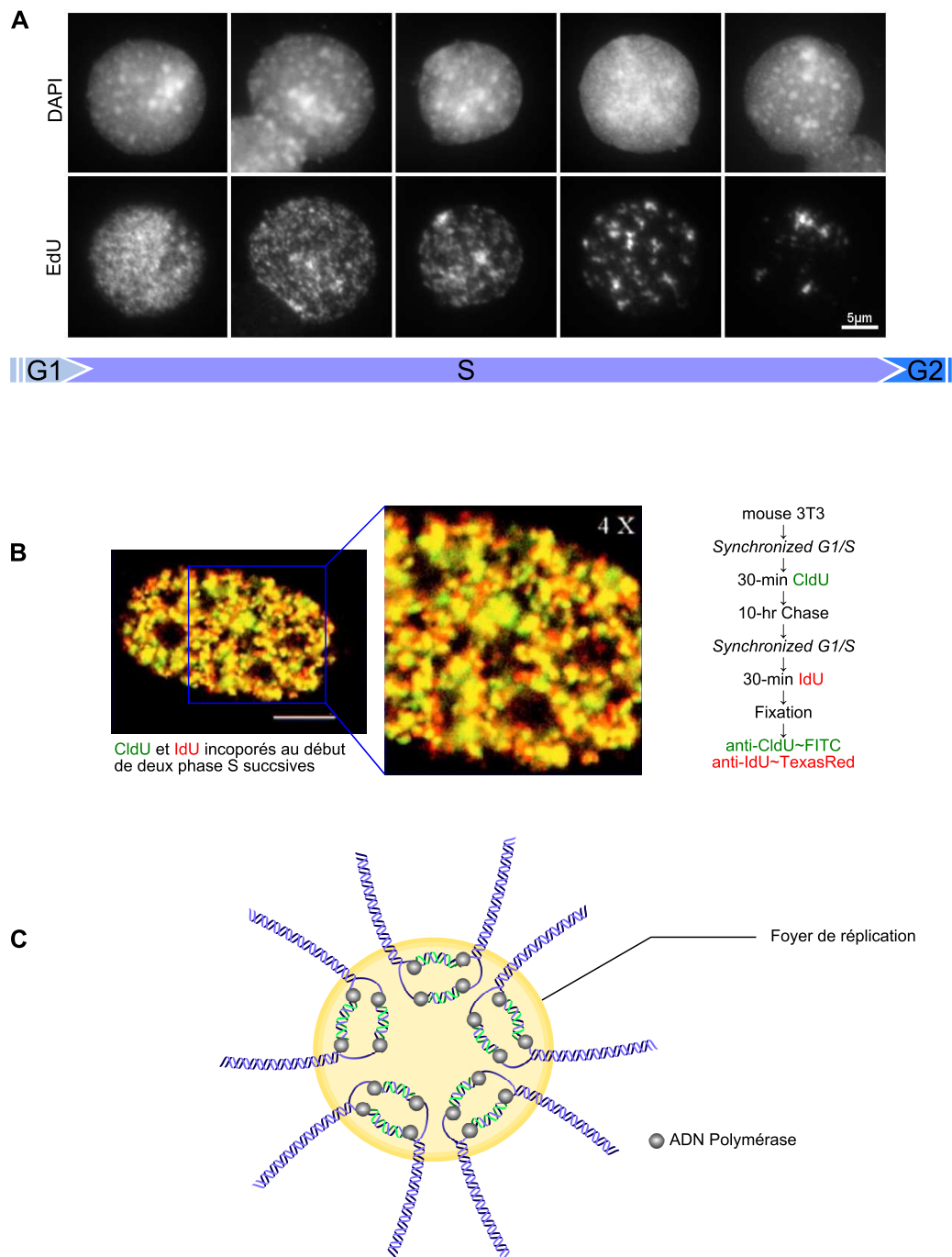


FIGURE 22 – Foyers de réplication au cours de la phase S.

A- Des lymphocytes humains ont été stimulés à la phytohemagglutinine pendant 68 h. 30 min avant la fixation au Carnoy (methanol 3 :1 acide acétique), le milieu de culture a été supplémenté par de l'EdU qui s'incorpore dans l'ADN en cours de synthèse.

B- Les foyers de réplication sont stables d'un cycle à l'autre. Du CldU et de l'IdU sont incorporés au début de 2 phases S successives. Le marquage jaune témoigne que les mêmes foyers de réplication sont utilisés d'un cycle à l'autre ([Ma et al., 1998], publié originellement dans *J Cell Biol* doi:10.1083/jcb.143.6.1415)

C- Usine de réplication. Le foyer jaune correspond aux sites d'incorporation de nucléotides halogénés visualisés par microscopie fluorescente. Selon le modèle classique, chaque foyer de début de phase S serait composé de l'agrégation de 5-6 séquences en cours de réplication

concentrent en de gros granules peu nombreux [Nakayasu and Berezney, 1989; O'Keefe et al., 1992; Sparvoli et al., 1994] (figure 22A). D'après la disposition des foyers de réplication, la phase S peut être subdivisée en 3 à 5 étapes avec chacune un agencement des foyers de réplication qui lui est propre. Des études par microscopie électronique [Koberna et al., 2005] ou par microscopie optique super-résolutive [Baddeley et al., 2010] ont permis de préciser la structure des gros foyers de réplication observés en fin de phase S : ils correspondraient plutôt à des conglomerats de petits foyers ayant une taille similaire à celle des foyers de début de phase S.

De nombreux acteurs moléculaires de la réplication (PCNA, RPA, ADN-polymérases α et ϵ , ADN ligase et topoisomérase II) colocalisent avec ces foyers de réplication [Frouin et al., 2003]. Pour chaque foyer, l'ensemble de la machinerie enzymatique est donc présente. L'association entre la protéine PCNA et les foyers de réplication est telle que la protéine fusion PCNA-GFP montre une parfaite colocalisation avec les sites d'incorporation de BrdU. L'agencement des spots de PCNA-GFP reproduit en outre les changements de disposition des foyers de réplication observés tout au long de la phase S [Leonhardt et al., 2000].

L'agencement des foyers de réplication étant différent en début et en fin de phase S, ceci soulève la question de la dynamique des foyers de réplication. Pour l'aborder, des cellules ont été marquées en début de phase S par du CldU, puis, après une chasse d'une durée variable, par de l'IdU. Ce double marquage montre que pour une chasse de 5 min, les deux marquages coïncident, mais qu'ils se séparent pour une chasse de plus de 30 min [Ma et al., 1998]. Le temps de vie d'un foyer de réplication en début de phase S a ainsi été évalué à ~45 min. L'observation de foyers différents après 45 min de chasse peut s'expliquer de deux manières différentes : soit les foyers sont mobiles et sont décalés 45 min plus tard, soit de nouveaux foyers apparaissent tandis que d'autres disparaissent. L'observation par time-lapse de la protéine fusion PCNA-GFP a permis de discriminer ces deux possibilités puisque que les spots PCNA-GFP sont immobiles. Les changements d'agencement des foyers de réplication au cours de la phase S résulteraient donc de l'assemblage de nouveaux foyers, et du désassemblage d'anciens [Leonhardt et al., 2000].

L'incorporation de nucléotides modifiés montre que la réplication de l'ADN se déroule en des sites discrets du noyau, qui présentent un agencement et des tailles variables au cours de la phase S. Ces foyers contiennent la machinerie enzymatique nécessaire à la réplication. Ils s'assemblent et se désassemblent au cours de la phase S.

b- Les foyers correspondent à des usines de réplication

Les cellules 3T3 possèdent, à un instant donné en début de phase S, en moyenne 1000 foyers de réplication dont le temps de vie est d'environ 45 min, et la phase S de ces cellules durent en moyenne 8 h. On peut alors estimer qu'il devrait se former $\approx 10\,000$ foyers de réplication tout au long de la phase S. Or les cellules 3T3 (polyplôides) possèdent environ 10 Gb d'ADN, soit 1 Mb par foyer de réplication. La vitesse des fourches (1,4 kb/min) ne permet pas de répliquer 1 Mb en 45 min. Pour répliquer une telle quantité d'ADN en 45 min, il faut imaginer le fonctionnement simultané de 10-14 fourches. Avec ce calcul approximatif, Ma et al. [1998] ont donc proposé que chaque foyer de réplication serait composé de 5-7 réplicons fonctionnant conjointement (figure 22C). Chaque foyer corres-

2.3- Structuration de la chromatine associée à la réplication

pondrait donc à une usine de réplication dans laquelle les réplicons regroupés partageraient les mêmes acteurs enzymatiques [Berezney et al., 2000; Berezney, 2002].

L'analyse de l'agencement des foyers de réplication après une longue chasse montre que cette organisation en foyers persiste au long du cycle cellulaire, puisque des spots BrdU sont observés aussi bien en phase G2, dans la phase G1 suivante [Jackson and Pombo, 1998] ou 44 h après le marquage [Sparvoli et al., 1994]. L'analyse du marquage BrdU après une chasse de 5 jours montre en outre que les foyers conservent la même taille et surtout la même intensité suggérant que les réplicons marqués au début de la phase S restent agglutinés sur plusieurs générations cellulaires [Jackson and Pombo, 1998].

De plus, les foyers de réplication de début et de fin de phase S, non seulement occupent des bandes distincts sur les chromosomes mitotiques, mais également des emplacements différents dans l'espace nucléaire avec très peu de chevauchement [Ma et al., 1998; Zink et al., 1998]. Cette ségrégation spatiale des régions précoces et tardives est maintenues pendant au moins 2 semaines [Zink et al., 1998], soulignant la stabilité de l'organisation nucléaire.

Enfin, ce sont les mêmes foyers de réplication qui sont activés au début de deux phase S successives (voir figure 22B), démontrant l'étonnante stabilité du programme spatio-temporel de la réplication [Ma et al., 1998; Dimitrova and Gilbert, 1999; Sadoni et al., 2004].

Les foyers de réplication sont constitués de plusieurs fourches se répliquant conjointement, sont stables sur plusieurs générations cellulaires, et sont organisés de telles sortes à ségréger les régions précoces des régions tardives. Ils représentent donc un niveau fondamental d'organisation tridimensionnelle de la chromatine [Berezney et al., 2000; Berezney, 2002]. Avec leurs 5-7 réplicons, ils constituent ainsi des unités structurales stables du noyau.

c- Fonctionnement de l'usine de réplication

Cette organisation en usine de réplication est compatible avec les modèles où les polymérases seraient fixes et où l'ADN serait mouvant. Ce dernier passerait dans l'usine au cours de sa réplication et serait ensuite extrudé (pour une revue, voir Cook [1999]). L'idée de complexes enzymatiques fixes est en accord avec (i) la localisation des foyers de réplication sur la matrice nucléaire [Radichev et al., 2005], (ii) la localisation de l'ADN naissant sur cette même matrice [Vaughn et al., 1990], et (iii) le recrutement des ORI à la matrice avant leur activation en phase G1 [Djeliova et al., 2001b] (voir aussi page 81).

Ce passage d'ADN à travers l'usine de réplication a été visualisé par time-lapse chez la levure. En cumulant l'insertion d'opérateur Tet (TetO) dans une région répliquée environ 20 min après l'activation de l'ORI la plus proche [Raghuraman et al., 2001], l'expression de TetR-CFP pour visualiser la localisation de TetO, et l'expression de la polymérase α couplée à la GFP pour visualiser les foyers de réplication, Kitamura et al. [2006] ont pu quantifier la coordination entre réplication du locus (augmentation de la fluorescence d'un facteur 2) et colocalisation avec le foyer de réplication. Ils ont ainsi remarqué la séquence suivante :

- avant réplication du locus (fluorescence = 1), le locus TetO ne colocalise pas avec les foyers de réplication
- pendant sa réplication (la fluorescence passe de 1 à 2), le locus TetO colocalise avec les foyers de réplication

- après réplication (fluorescence = 2), TetO se localise à nouveau en dehors de foyers de réplication

Cette séquence d'événements est en accord avec le passage de l'ADN dans l'usine pour sa réplication, suivi de son extrusion une fois la duplication effectuée.

En plus d'analyser ce passage d'ADN à travers un foyer de réplication, [Kitamura et al. \[2006\]](#) ont pu tester l'association de plusieurs fourches au sein d'une même usine de réplication. Pour cela, ils ont également inséré des opérateurs Lac (LacO) de telle sorte que TetO et LacO soient localisés de part et d'autre de la même origine et tous les deux répliqués de manière synchrone 20 min après l'activation de l'origine centrale (figure 23A). En exprimant TetR-CFP et LacI-GFP, ces auteurs ont pu mesurer la distance les séparant au cours de la phase S. Si les deux fourches émanant de l'ORI centrale sont localisées dans la même usine de réplication, la distance entre LacO et TetO devrait être petite lorsque les deux loci sont en train de se répliquer. Le résultat de cette expérience montre que (i) TetO et LacO se répliquent bien au même moment (la fluorescence de TetR-CFP et LacI-GFP double simultanément), et (ii) que la distance séparant les deux loci est divisée par 6 pendant 3 min juste avant leur duplication, suggérant une colocalisation transitoire des deux loci au même foyer de réplication (figure 23B). Ce résultat démontre donc que deux fourches sœurs émanant de la même origine de réplication se trouvent au sein de la même usine de réplication (figure 23C).

Les foyers de réplication, qui constituent un niveau d'organisation de la chromatine stable sur plusieurs générations cellulaires, contiennent, en phase S, plusieurs fourches de réplication fonctionnant conjointement. La machinerie de réplication y est fixe et l'ADN mouvant à travers l'usine de réplication. Cette organisation autour des foyers de réplication implique la formation de contacts préférentiels entre certaines portions de chromatine, notamment entre celles se répliquant de manière synchrone. Cette organisation de la chromatine étant stable (c'est-à-dire présente en phase G1 et G2, mais aussi dans les cycles ultérieurs), on peut supposer que ces contacts préférentiels peuvent être détectés par des méthodologies de la famille 3C (voir page 37)

2.3.3- Contacts chromatinien et réplication

a- Chez la levure – interactions entre ARS

Des contacts entre différentes portions de chromatine peuvent être détectés par les méthodologies de la famille 3C (voir page 37). Les ARS étant connues chez la levure, les données Hi-C de *S. cerevisiae* permettent de tester s'ils établissent des interactions préférentielles. Parmi l'ensemble des interactions détectées, un enrichissement a été observé entre origines de réplication précoces mais pas entre origines tardives [[Duan et al., 2010](#)]. Cette différence de comportement selon le timing de réplication pourrait s'expliquer par la persistance des interactions entre régions précoces en phase G1 et G2, alors que les interactions entre régions tardives pourraient ne s'établir que pendant la phase S, lors de la réplication des loci en question.

Parmi les origines interagissant, certaines seraient activées par les facteurs Fkh (Forkhead transcription factors) dont le rôle dans le déroulement de la réplication commence à être étudié [[Knott et al., 2012](#)]. En utilisant l'une de ces origines

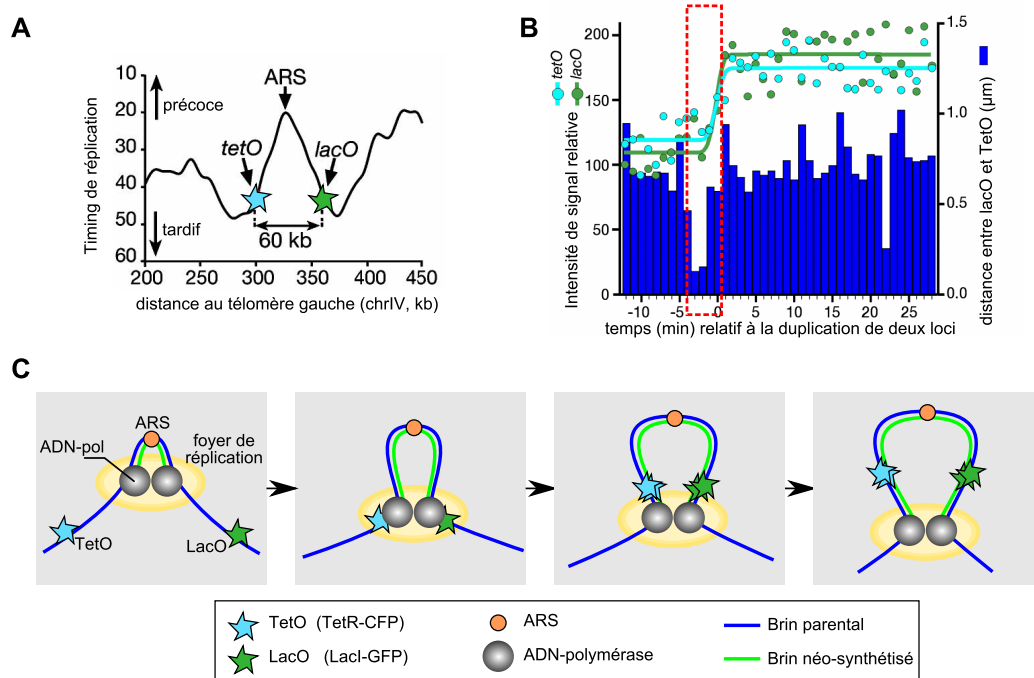


FIGURE 23 – Coordination de deux fourches de réplication au sein du même foyer de réplication. A- Afin de visualiser la progression de deux fourches sœurs, des opérateurs Lac (LacO) et Tet (TetO) ont été insérés chez la levure de part et d'autre d'une origine de réplication (ARS). Ces deux opérateurs sont visualisés grâce à l'expression des protéines LacI-GFP et TetR-CFP. B- TetO et LacO colocalisent juste avant leur réplication. La fluorescence des spots GFP et CFP est mesurée au cours du temps ; le doublement de la fluorescence correspond au moment où ces loci sont dupliqués par le passage de la fourche de réplication. La distance entre les deux spots est également mesurée : juste avant leur réplication, TetO et lacO colocalisent. C- Modèle explicatif : les deux fourches sœurs émanant de l'origine (ARS) centrale sont fixées dans la même usine de réplication, conduisant à la colocalisation de TetO et LacO lorsqu'ils sont répliqués. Ces figures sont issues et adaptées de Kitamura, E., Blow, J. J., and Tanaka, T. U. (2006). Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell*, 125(7):1297–1308, copyright 2006, avec l'autorisation d'Elsevier.

comme point de vue pour une expérience 4C, il été démontré qu'elle interagit effectivement avec d'autres origines, elles-mêmes régulées par Fkh. Ces interactions pourraient être établies par les facteurs Fkh directement puisque leur knock-out supprime ces interactions.

b- Chez l'homme – couplage du timing de réplication et des données Hi-C

Chez les mammifères, les origines de réplication sont mal connues. De fait, tester des interactions entre origines de réplication n'est pas directement possible. Néanmoins, de nombreuses données de timing de réplication sont disponibles. Couplées au Hi-C effectué chez l'homme [Lieberman-Aiden et al., 2009], elles permettent de tester si des portions de chromatine répliquées de manière synchrone établissent des interactions.

Dans le Hi-C, les séquences obtenues sont chimériques, la moitié de gauche correspond au partenaire 1 et la moitié de droite au partenaire 2 (figure 10). En comptant le nombre fois où les deux partenaires ont le même timing de réplication et le nombre de fois où les deux partenaires ont des timings différents, on peut donc tester si les régions se répliquant de manière synchrone interagissent

plus que des régions asynchrones. Pour cela, [Yaffe et al. \[2010\]](#) se sont focalisés sur les interactions inter-chromosomiques, de manière à s'affranchir de toutes les interactions locales qui sont plus abondantes et, puisqu'elles sont locales, ont plus de chance de s'effectuer entre régions possédant le même timing de réplication. Ces précautions prises, les contacts inter-chromosomiques apparaissent plus fréquents entre deux régions précoces, qu'entre une région précoce et une autre tardive. Un résultat identique a été obtenu en considérant les interactions entre deux régions tardives, démontrant donc que les portions de chromatine répliquées de manière synchrone interagissent préférentiellement.

En utilisant une approche légèrement différente, [Ryba et al. \[2010\]](#) aboutissent au même résultat. Comme évoqué (voir page 44 et suivantes), l'un des principaux résultats du Hi-C est la segmentation du génome en deux compartiments (A et B), montrant une exclusion mutuelle. Pour aboutir à ce résultat, les données Hi-C ont été analysées par composante principale et un vecteur explicatif (vecteur propre) a été trouvé. La valeur du vecteur propre pour un locus donné détermine son appartenance à l'un des deux compartiments. [Ryba et al. \[2010\]](#) ont alors comparé ce vecteur propre des données Hi-C avec les données de timing de réplication et ont observé une corrélation extrêmement forte entre les deux signaux, plus forte que pour n'importe quelle modification d'histones. Ce résultat démontre donc que les régions précoces appartiennent plutôt au compartiment A et interagissent donc plutôt avec d'autres membres du compartiment A (donc avec d'autres régions précoces). L'inverse est également vrai : les régions tardives appartiennent plutôt au compartiment B et interagissent donc avec d'autres membres du compartiment B (soit avec d'autres régions tardives). Ces travaux démontrent donc que les portions de chromatine se répliquant de manière synchrone interagissent ensemble.

Une dernière étude aborde la même question. Grâce à l'utilisation de données d'altérations somatiques du nombre de copies (événements d'amplifications ou de délétions) observées dans de nombreux cancers, [De and Michor \[2011\]](#) ont remarqué que les bordures de ces altérations du nombre de copies résident plus souvent entre deux régions interagissant en Hi-C qu'attendu. Or, ces bordures d'altération du nombre de copies ont également plus de chance de résider dans des régions génomiques ayant, chez le sujet sain, le même timing de réplication. En combinant ces deux observations, ces auteurs proposent que la proximité spatiale entre deux loci se répliquant simultanément puisse être un bon prédicteur des patrons de mutations observés dans les cancers.

Pris dans leur ensemble, ces trois études aboutissent à une conclusion commune : les régions de timing opposé ségrègent dans l'espace nucléaire, permettant une interaction entre régions précoces d'un côté, et entre régions tardives de l'autre.

c- Limites de ces études chez l'homme

Le jeu de données Hi-C, rappelons-le, résume l'ensemble des interactions deux-à-deux présentes dans les cellules Gm06990 et K562. Il est donc possible d'extraire de ce jeu de données les interactions impliquant un locus particulier (un équivalent du 4C peut être extrait des données Hi-C). En représentant les interactions de 4 régions tardives (L1, L2, L3, L4, voir figure 24A), le long de 20 Mb du chromosome 5, il est difficile de conclure que L4 interagit avec ses voisines tardives (même si quelques séquences sont capturées en face de L2 et peut-être de L1). Toutes les réciproques sont vraies : sur les trois autres "4C" extraits du Hi-

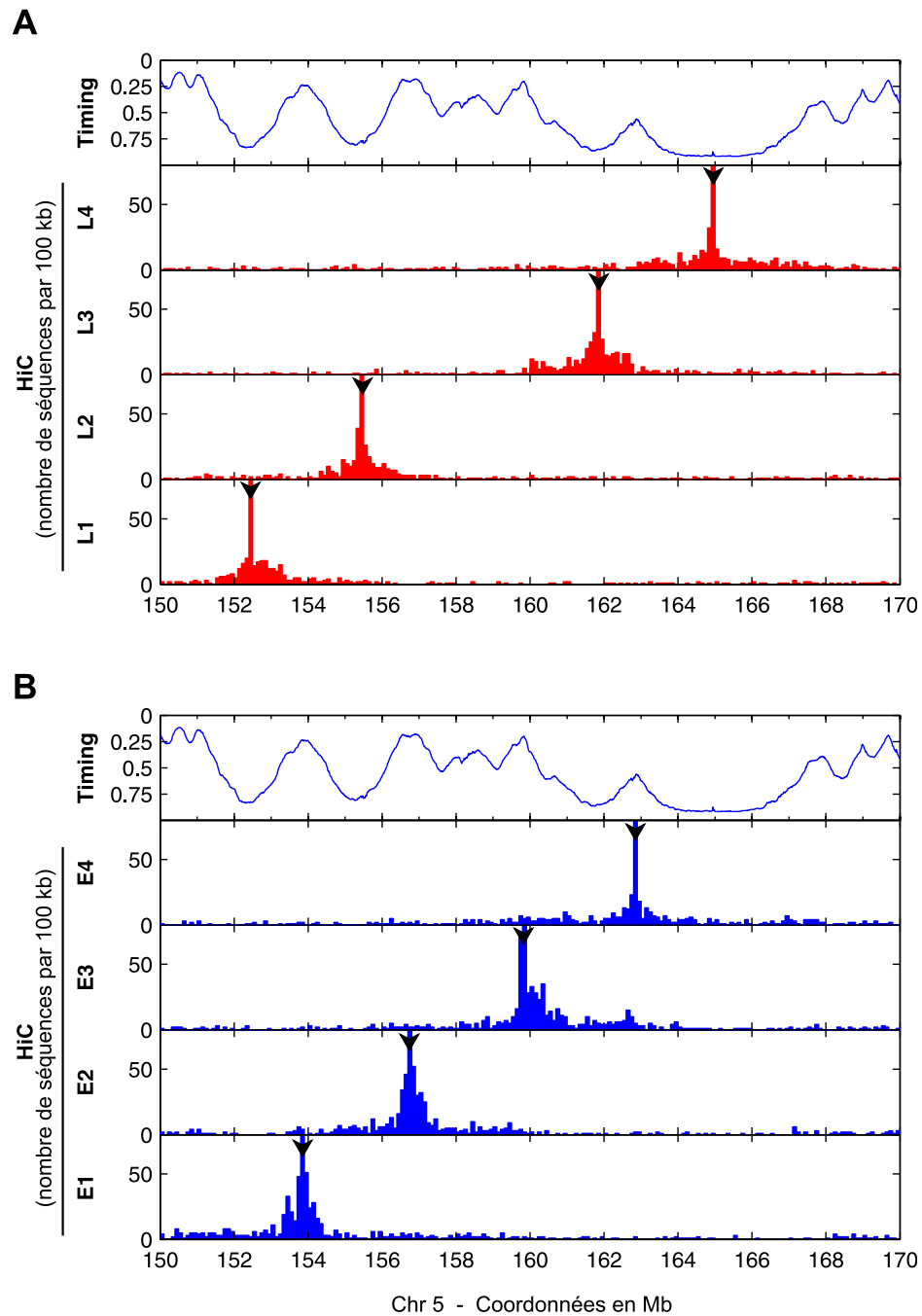


FIGURE 24 – Hi-C et réplication – les limites. Les données Hi-C contiennent l'ensemble des interactions deux à deux. On peut donc en extraire les interactions impliquant un locus particulier. Ici, sur une portion de 20 Mb du chromosome 5 humain, on extrait les interactions de 4 loci répliqués en fin de phase S (A : L1, ..., L4) et de 4 loci répliqués en début de phase S (B : E1, ..., E4). La position de ces points de vue est indiquée par les têtes de flèches. Le timing de réplication est représenté dans les premiers panneaux. L'axe des ordonnées a été ajusté de manière à bien visualiser les régions où le nombre de lectures par 100 kb est faible. Hi-C et timing proviennent des cellules lymphoblastoïdes Gm06990 [Lieberman-Aiden et al., 2009; Hansen et al., 2010].

C, on ne voit pas d'interactions flagrantes entre régions tardives voisines. Notez également la faible ordonnée : les barres autour de 152 Mb dans le panneau L4 correspondent à quelques lectures (1 à 3) dans une fenêtre de 100 kb.

Faisons maintenant de même avec des régions répliquées en début de phase S (figure 24B). Dans ce cas, on peut noter quelques interactions entre voisins précoces. Par exemple, E3 présente un rebond en face de E4, de même pour E2 en face de E1. Néanmoins, passé le voisin immédiat, les interactions entre régions précoces, si elles existent, sont difficiles, voire impossibles, à visualiser.

Les trois études évoquées précédemment qui comparaient Hi-C et timing de réplication ont pourtant réussi à détecter des interactions s'effectuant entre régions répliquées de manière synchrone. Ce résultat est juste, mais dérive principalement d'un effet de moyenne sur l'ensemble du génome :

- Le vecteur propre principal utilisé par Ryba et al. [2010] est établi par l'analyse en composante principale en prenant en compte l'ensemble des interactions d'une région donnée. Par conséquent, comparer le timing de réplication au vecteur propre principal du Hi-C revient à prendre en compte l'ensemble des partenaires identifiés dans le génome.
- En comptant, pour les *trans*-interactions, le nombre de fois où les deux partenaires résident dans des régions à timing similaire, Yaffe et al. [2010] introduisent par construction un effet de moyenne important. Dans cette approche, l'ensemble de toutes les régions précoces et de leur patron d'interaction à chacune sont résumés en une seule valeur.
- Un effet de moyenne existe également dans l'étude de De and Michor [2011] dans laquelle ce sont les bordures de plus de 300 000 altérations somatiques du nombre de copies qui sont résumées en une seule valeur.

Pour résumer, dans ces trois études, les partenaires réels d'une région précoce ou d'une région tardive ne sont pas directement visualisables, vraisemblablement à cause du manque de résolution des données Hi-C. En effet, rappelons que la complexité d'une analyse de l'ensemble des interactions 2 à 2 d'une cellule est de 10^{12} (voir page 40), soit 10 000 fois plus que le nombre de lectures séquencées (de l'ordre de 10^7 à 10^8). Toutes les combinaisons ne sont donc pas accessibles. Pour un point de vue donné (1 fragment de restriction sur les 10^6 existants), il n'y aura en moyenne que $\frac{10^8}{10^6} = 100$ lectures pour identifier ses partenaires. La comparaison des données Hi-C avec la réplication présente donc un certain nombre de limites, et, entre autre, ne permet pas la visualisation directe des partenaires d'un locus précoce ou d'un locus tardif.

LES QUESTIONS ABORDÉES EXPÉRIMENTALEMENT DANS CE MANUSCRIT

Projet de thèse

La comparaison des données Hi-C et du timing de réplication laisse des questions en suspens.

1. Quels sont, s'ils existent, les partenaires préférentiels d'une région répliquée en début (réciproquement en fin) de phase S ?
2. En excluant tout effet de moyenne, peut-on voir une interaction entre deux régions voisines répliquées au même moment dans la phase S ?
3. En accord avec l'observation que domaines fonctionnels et domaines structuraux se recoupent, peut-on voir des frontières physiques associées aux domaines fonctionnels de timing ? Où se localisent-elles ? Quelle est la signature chromatinienne de ces frontières ? Permettent-elles de renforcer l'une des deux segmentations du timing proposées (segmentation à timing constant avec des domaines plats, ou domaines en forme de U, voir page 75) ?
4. Ces interactions entre régions répliquées au même moment de la phase S peuvent-elles être visualisées (sans effet de moyenne) sur une dizaine de Mb (les quelques régions synchrones alentour), voire sur plusieurs dizaines de Mb, ou la centaine de Mb (chromosome entier) ?
5. Les interactions entre régions précoces sont-elles maintenues entre différents chromosomes (*trans*-interactions) ?
6. Les interactions s'établissent-elles uniquement lors de la réplication ? L'observation des foyers de réplication (voir page 82 et suivantes) montre que les foyers de réplication sont maintenus sur plusieurs générations cellulaires, y compris en phase G1. Ce résultat suggère que les contacts établis persistent, y compris hors de la phase S. Visualiser les interactions de régions précoces et tardives sur des cellules en cycle et des cellules hors-cycle permettrait d'apporter une première preuve de leur existence en dehors de la phase S.

Les études cytologiques ne peuvent pas aborder la plupart de ces questions d'une manière directe et catégorique. Par contre, la mise en place d'une expérimentation 4C devrait permettre d'y apporter une réponse. En effet, à la différence du Hi-C, le 4C fournit une image nettement plus détaillée des contacts chromatinien d'un locus sélectionné [de Wit and de Laat, 2012]. Nous proposons donc d'utiliser plusieurs points de vue 4C (certains répliqués en début de phase S, et d'autres répliqués en fin de phase S) pour déterminer leurs principaux partenaires. Nous proposons également d'effectuer cette étude chez des lymphocytes cyclants et non-cyclants (bloqués en phase G0) pour apporter de premiers indices quant au maintien de contacts chromatinien en lien avec la réplication dans des cellules hors cycle.

Partie B

Résultats

B

Mise en place d'une stratégie 4C-Seq

3.1	Choix de la région et du type cellulaire	97
3.1.1	Sélection d'une région à timing conservé et contrasté (chr5 :150-170Mb)	97
	<i>Cette région montre un timing similaire dans différentes cellules</i>	97
	<i>Cette région montre un timing similaire d'un laboratoire à l'autre</i>	98
	<i>Cette région montre un timing contrasté</i>	98
3.1.2	Choix du type cellulaire	99
3.2	De la librairie 4C au séquençage haut-débit	100
3.2.1	Préparation des librairies 4C	100
	<i>Un protocole adapté de celui du groupe de Wouter de Laat</i>	100
	<i>Contrôle de la digestion sur chromatine cross-linkée</i>	101
	<i>Contrôle de la ligation et de la digestion par Hin1II</i>	103
3.2.2	Positionnement et conception des oligonucléotides 4C	104
3.2.3	Profils des PCR 4C	104
3.2.4	Des PCR 4C-Seq au séquençage haut-débit	107
	<i>Réplicats et profils de migration des PCR 4C-Seq</i>	107
	<i>Multiplexer les PCR 4C-Seq</i>	108
	<i>Caractéristiques du séquençage haut-débit</i>	110
3.3	Bioinformatique et traitement du séquençage haut-débit	110
3.3.1	Dé-multiplexer les lectures 4C	110
3.3.2	Alignement sur le génome	112
3.3.3	Filtrage des lectures en fonction de leur alignement	115

Afin d'étudier de manière directe les liens entre organisation tridimensionnelle de la chromatine et réplication, nous avons mis en place une stratégie 4C couplée à un séquençage haut-débit (4C-Seq). L'expérimentation 4C (voir page 37) donne une vision statistique des partenaires d'un locus sélectionné, appelé "point de vue". En utilisant divers points de vue, nous proposons d'identifier les partenaires de loci répliqués en début de phase S et de les comparer à ceux de loci répliqués en fin de phase S. L'analyse des contacts préférentiels de chacun fournira une première description des liens entre réplication et organisation de la chromatine.

Dans ce chapitre, qui mêlera volontairement méthodologie et résultats / contrôles intermédiaires, je détaillerai la démarche expérimentale mise en place. Les profils 4C seront quand à eux exploités dans le chapitre suivant. Nous suivrons une démarche en 3 étapes :

- 1 nous sélectionnerons une région du génome sur laquelle sera conduite notre étude et expliciterons les raisons ayant motivé ce choix ;
- 2 puis nous insisterons sur les étapes expérimentales du protocole et les contrôles ;
- 3 enfin, nous décrirons l'approche bioinformatique.

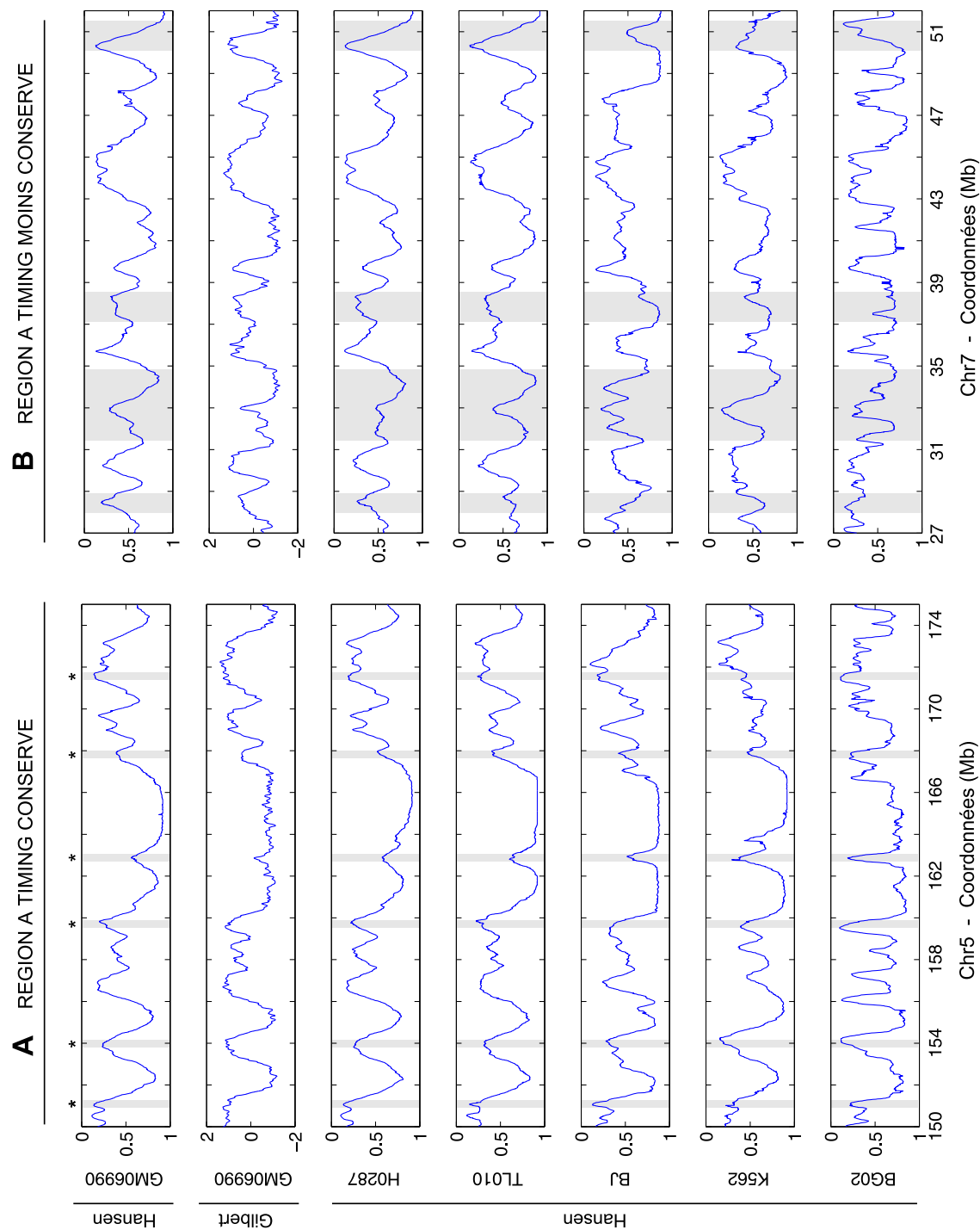


FIGURE 25 – Choix d'une région à timing de réplication conservé. Certaines régions (A) ont un timing conservé d'un type cellulaire à l'autre alors que d'autres (B) non. Le panneau 2 en partant du haut provient du site <http://www.replicationdomain.com> (D.M. Gilbert), les autres proviennent de Hansen et al. [2010]. De haut en bas, des cellules lymphoblastoïdes (GM06990 et H0287), des cellules lymphoblastoïdes avec le syndrome de l'X fragile (TL010), des fibroblastes normaux (BJ), une lignée cellulaire à propriétés érythrocytaires (K562), et des cellules souches (BG02). En A, les ombres verticales et l'astérisque indiquent des pics de timing conservés dans au moins 5 des 6 panneaux de l'analyse Hansen et al. En B, les ombres verticales indiquent des régions de timing différent.

3.1- Choix de la région et du type cellulaire

3.1.1- Sélection d'une région à timing conservé et contrasté (chr5:150-170Mb)

Le timing de réplication utilisé comme référence dans ce manuscrit est celui établi par [Hansen et al. \[2010\]](#), car il couvre tout le génome humain, il est établi avec une haute résolution temporelle (phase S sub-divisée en 6). De plus, il a été analysé dans notre laboratoire pour y détecter des U-domaines de timing (voir page 75 et suivantes ainsi que [\[Baker et al., 2012\]](#)).

Pour notre étude, nous avons sélectionné la région du chromosome 5 humain comprise entre 150 et 170 Mb pour plusieurs raisons exposées dans cette sous-section.

a- Cette région montre un timing similaire dans différentes cellules

Pour décrire les profils de timing, on utilisera dans ce manuscrit une terminologie qu'il vaut mieux introduire dès maintenant : un *pic de timing* correspond à un maximum local dans le profil, c'est-à-dire à une région répliquée plus précocement que ces voisins immédiats ; une *vallée de timing* est l'opposé, à savoir une région répliquée plus tardivement que les régions adjacentes.

La comparaison du timing de réplication de différents types cellulaires montre qu'il est globalement conservé sur ces 20 Mb du chromosome 5 (figure 25A).

- En effet, 3 types de cellules lymphoblastoïdes (Gm06990, H087 et TL010) ont globalement le même timing avec seulement quelques subtiles différences :
 - à 158,5 Mb, on observe une unique bosse de timing chez les H0287, 2 bosses chez les TL010, et une situation intermédiaire chez les Gm06990 ;
 - à 164 Mb, on observe un rebond dans le timing des H0287, mais pas dans les deux autres types de cellules lymphoblastoïdes.
- Pour des types cellulaires différents, les divergences de timing sont légèrement plus marquées, même si elles restent minimales :
 - à 164 Mb, on observe un franc pic de timing dans les cellules K562 qui n'est pas observé dans les autres types cellulaires analysés (voir chapitre 5) ;
 - les 3 pics de timing de la région [157 – 160 Mb] présentent une forme différente dans les cellules lymphoblastoïdes, fibroblastes (BJ) et K562 ;
 - les cellules souches BG02 présentent un timing relativement éloigné des cellules différenciées, en accord avec l'observation qu'entre 20 et 45% du génome change de timing lors de la différenciation [\[Hiratani et al., 2008; Desprat et al., 2009\]](#).
- Cette région présente donc un timing de réplication globalement partagé par plusieurs types cellulaires (les astérisques, figure 25A, indiquent des pics communs à la plupart des panneaux). Cette très forte conservation est une spécificité de cette région (en comparaison, la région [27 – 52 Mb] du chr 7 montre un timing de réplication nettement moins conservé d'un type cellulaire à l'autre, figure 25B).

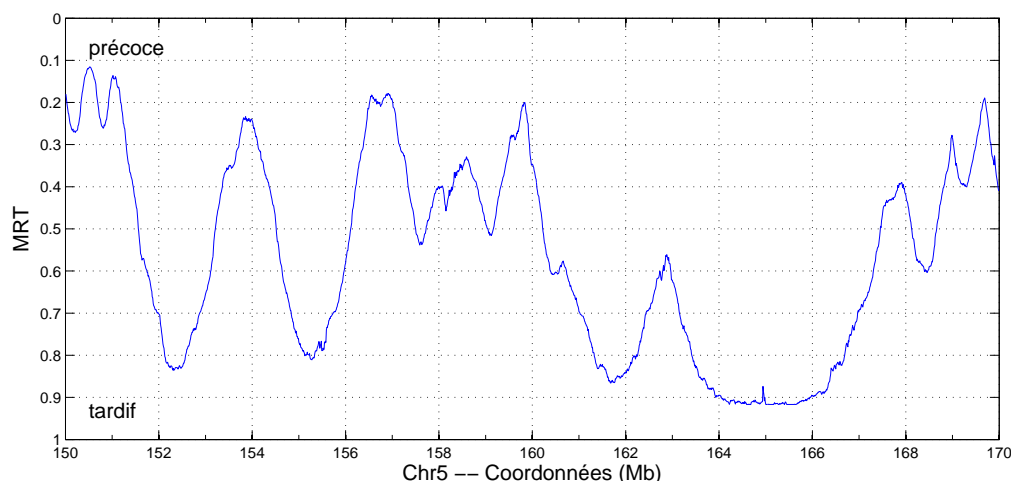


FIGURE 26 – Timing de réplication (MRT) de la région chr5:150–170 Mb. Le signal brut provient de Hansen et al. [2010]. Il a été post-traité par Baker et al. [2012] (voir page 75)

b- Cette région montre un timing similaire d'un laboratoire à l'autre

Le timing de réplication des cellules Gm06990 a également été déterminé par le groupe de D.M Gilbert ([Ryba et al., 2010] et <http://www.replicationdomain.com>). Au niveau de la région du chr 5 choisie, leur timing s'avère très proche de celui établi par Hansen et al. [2010]. Même si l'on ne peut pas exclure que les mêmes artéfacts/biais surviennent dans les deux laboratoires, l'excellente reproductibilité d'un laboratoire à l'autre (figure 25A, panneaux 1 et 2) témoigne certainement de l'exactitude du timing établi.

c- Cette région montre un timing contrasté

La région chr5:150-170Mb possède un timing fluctuant (figure 26), avec des loci répliqués en tout début de phase S (abscisse proche de zéro, par exemple à 154 Mb ou 162,8 Mb), et d'autres répliqués en toute fin phase S (abscisse proche de 1, par exemple à 152,3 Mb ou 165 Mb). Des situations intermédiaires existent également, par exemple :

- un pic de timing répliqué au milieu de la phase S (à 162.7 Mb, timing = 0.55) ;
- des vallées de timing pas très tardives (par exemple à 159 Mb).

Dans ces 20 Mb du chr 5, des points de vue 4C peuvent donc être placés sur des loci précoces ou tardifs, mais aussi sur ces situations intermédiaires. Un pic de timing répliqué en milieu de phase S nous permettra d'évaluer si c'est la valeur précise du timing qui détermine les contacts à longue-distance ou si c'est le fait d'être plus précoce que ses voisins (valeur relative). Enfin, les régions précoces de cette région du chr 5 sont très fines, permettant le placement sans ambiguïté d'amorces 4C sur les pics de timing.

La région du chr 5 choisie présente un timing de réplication conservé entre plusieurs types cellulaires. Il est contrasté et montre des changements brusques au niveau des régions précoces, permettant de positionner précisément les points de vue 4C.

3.1.2- Choix du type cellulaire

- Nous avons choisi de mener notre étude sur des cellules lymphoblastoïdes pour plusieurs raisons :
 - le timing de réplication est établi dans les cellules lymphoblastoïdes Gm06990 par au moins deux groupes différents [Hansen et al., 2010; Ryba et al., 2010];
 - de nombreuses marques épigénétiques et facteurs chromatinien ont été étudiés dans ces cellules, et de nombreux ChIP-chip ou ChIP-Seq sont disponibles, autorisant les méta-analyses.
- Notre étude a également été conduite sur des cellules mononucléées du sang périphérique (PBMC, constituées majoritairement de lymphocytes [Delves et al., 2006]) pour deux raisons. Premièrement, il s'agit de cellules normales, non tumorales, et non transformées par un virus. Deuxièmement, elles sont bloquées en phase G0 et ne se répliquent donc pas. La comparaison des PBMC et des cellules lymphoblastoïdes permettra d'aborder la question de la stabilité des contacts chromatinien dans des cellules en cycle et hors-cycle.
- Un 4C-Seq a également été effectué sur des PBMC stimulées à la phytohé-magglutinine, une lectine connue pour activer les lymphocytes et les sortir de dormance.
- Ces trois types cellulaires étant proches, le timing de réplication des cellules lymphoblastoïdes Gm06990 a été utilisé comme référence pour cha-cune.

Les liens entre organisation à grande échelle de la chromatine et timing de réplication ont été étudiés sur des lymphocytes et dérivés de lymphocytes.

Protocole : Culture cellulaire

Toutes les cellules sont cultivées à 37°C avec 5% CO₂ dans du RPMI supplémenté avec 10% sérum, glutamax et pénicilline/streptomycine. Les PBMC sont isolées du sang de donneurs sains par centrifugation sur Ficoll-Paque PLUS (GE Healthcare) avec les tubes Leucosep®(Greiner bio-one) en suivant les instructions des fabricants. Les PBMC sont stimulées par ajout dans le milieu de culture de 1,5% phytohé-magglutinine (PHA) pendant 68 h. 10⁶ PBMC stimulées ont été fixées à l'éthanol et traitées à l'iodure de propidium (20µg/mL, 30 min, 37°C) pour une analyse du cycle par FACS (LSRII, FACSDiva™, BD Biosciences).

3.2- De la librairie 4C au séquençage haut-débit

Dans cette section, nous aborderons les étapes expérimentales en insistant sur les contrôles intermédiaires.

3.2.1- Préparation des librairies 4C

a- Un protocole adapté de celui du groupe de Wouter de Laat

Le protocole 4C utilisé est adapté de celui du groupe de Wouter de Laat [Simonis et al., 2006]. Afin de constituer un document de travail, je vais détailler dans ce paragraphe chaque étape de la préparation des librairies 4C à partir de cellules lymphoblastoïdes ou de PBMC.

Protocole : Préparation des librairies 4C

Cross-link Les cultures de lymphoblastoïdes / PBMC sont centrifugées (5 min, 200 g), et resuspendues dans du PBS - 10% sérum à une concentration de 2.10^6 cellules par mL. Un volume de solution de fixation (4% formaldéhyde dans du PBS / 10% sérum, préparée extemporanément) est alors ajouté (incuber 10 min à température ambiante sur tubes roulants). La réaction est arrêtée par ajout de glycine 1 M froide (qsp. 125 mM), et les cellules centrifugées (8 min, 220 g) pour enlever le fixateur.

Lyse Les cellules lymphoblastoïdes ou les PBMC sont incubées 10 min dans du tampon de lyse (10 mM Tris pH7.5, 10 mM NaCl, 0.3% NP-40, 1X inhibiteur de protéase) à 2.10^6 cellules par mL, puis lysées à l'aide d'un homogénéisateur de type dounce (≈ 20 coups de pistons B). Le déroulement de la lyse peut-être suivi par coloration au Vert de méthyle - Pyronine. Les noyaux sont récupérés par centrifugation (5 min, 400 g, 4°C).

Digestion par HindIII Des aliquotes de $6.6.10^6$ cellules sont préparées dans des eppendorfs de 1.5 mL, avec 510 μ L de tampon de restriction 1.17X. La chromatine est extraite par ajout de 7,5 μ L de 20% SDS (1h, 37°C, sous agitation), puis le SDS est piégé par ajout de triton (75 μ L de 20% Triton X-100, 1h, 37°C, sous agitation). La chromatine est enfin digérée par 300U de HindIII (4h, 37°C) puis à nouveau par 400U de la même enzyme (O/N, 37°C). [contrôle de la digestion possible, figure 27].

Ligation La digestion est arrêtée par inactivation de l'enzyme de restriction (ajout de 40 μ L de 20% SDS, 1h, 37°C). La chromatine digérée est transférée dans un falcon de 50 mL dans lequel le volume est ajusté à 7mL avec du tampon de ligation sans ATP (40 mM Tris pH7.8, 10 mM MgCl₂, 5 mM DTT, 70 μ g/mL BSA final). Le SDS est piégé par ajout de 375 μ L de 20% Triton (1h, 37°C), puis sont ajoutés l'ATP (1 mM final) et la ligase (150U, incubation O/N, 16°C).

Purification Le cross-link est inversé par incubation à 65°C (O/N) en présence de 300 μ g de protéinase K. L'ARN dégradé (400 μ g de RNase), une extraction au phénol-chloroforme est effectuée. La phase aqueuse est ensuite

diluée par 2, puis précipitée à l'éthanol en présence de glycogène (une fois l'alcool ajouté, le falcon peut être incubé à -80°C pour faciliter la précipitation). L'ADN est resuspendu dans 200 µL de 10 mM Tris pH 7.4 [contrôle sur gel possible, figure 28].

Circularisation L'ADN purifié est digéré par 50U de Hin1II (O/N, 37°) dans un volume final de 500 µL, puis purifié par extraction au phénol-chloroforme et précipitation à l'éthanol [contrôle sur gel possible, figure 28]. Une ligation est ensuite effectuée (14 mL, 1X Tampon de ligation avec 1 mM ATP, 300U ligase).

Purification Le produit de ligation est purifié par extraction au phénol-chloroforme, la phase aqueuse est diluée 2 fois et l'ensemble précipité à l'éthanol (une fois l'alcool ajouté, le falcon peut être incubé à -80°C pour faciliter la précipitation). L'ensemble est repris dans 300 µL de 10 mM Tris pH7.4, puis purifié par chromatographie (Roche High Pure PCR Product Purification, 3 colonnes par 10⁶ cellules). L'ADN est quantifié par fluorimétrie (QuBit dsDNA BR).

Les étapes les plus critiques de ce protocole sont la fixation, la digestion par HindIII et la ligation. S'il est difficile de suivre le déroulement de la fixation, les étapes de digestions enzymatiques ainsi que la première ligation peuvent être aisément contrôlées pour s'assurer de leur bon fonctionnement.

b- Contrôle de la digestion sur chromatine cross-linkée

Pour capturer des interactions 4C, une digestion efficace de la chromatine cross-linkée est essentielle. Le bon déroulement de la digestion a été évalué sur gel d'agarose en comparant les profils de migration d'une aliquote avant et après digestion. Sur gel, une chromatine convenablement digérée conduit à une migration sous forme de "smear", avec une bande majoritaire en haut de la trainée [Louwers et al., 2009]. Les profils de migration après digestion et purification (voir protocole) présentés figure 27A sont conformes à ce qui est attendu. En particulier, le "smear" est bien visible et la bande majoritaire nettement moins intense que pour l'aliquote avant digestion.

Protocole : Contrôle du taux de digestion

Une aliquote de 10 µL avant et après la digestion HindIII est sauvée, ajustée à 100 µL avec du 10 mM Tris pH7.4, incubée à 65°C (1h) en présence de 100 µg de protéinase K, et une extraction au phénol-chloroforme est effectuée. 20 µL sont déposés sur gel 0.7%, le reste est précipité à l'éthanol. L'ADN purifié est utilisé comme matrice pour des PCR quantitatives à l'aide d'amorces chevauchant un site HindIII.

Pour rigoureusement quantifier le taux de digestion, il est possible de mesurer le taux de coupure de certains sites HindIII par PCR quantitative [Hagège et al., 2007]. A l'aide d'oligonucléotides situés de part et d'autre d'un site HindIII, nous avons donc quantifié le nombre de copies amplifiables après digestion de la chromatine. Cette valeur a été comparée à celle obtenue avec des amorces

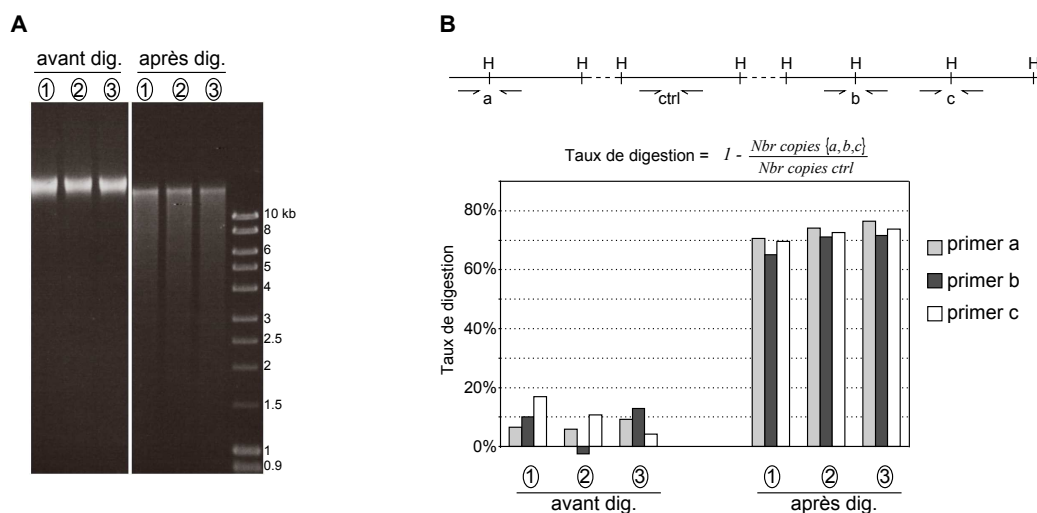


FIGURE 27 – Contrôle de la digestion sur chromatine cross-linkée de cellules lymphoblastoïdes. Pour capturer des interactions par 4C, il faut au préalable assurer une bonne digestion enzymatique par HindIII.

A- Évaluation du taux de digestion sur gel. Le cross-link est inversé sur une aliquote (protéinase K, 1 h, 65°C), une extraction au phénol-chloroforme est effectuée, puis l'ADN est déposé sur gel d'agarose 0.7%. On compare les profils avant digestion et après digestion. Après digestion, un "smear" est observé avec une bande majoritaire en haut du "smear". Les trois pistes correspondent à trois échantillons de cellules lymphoblastoïdes traités simultanément.

B- Évaluation du taux de digestion par PCR quantitative. L'ADN des aliquotes avant et après digestion est précipité, et une PCR quantitative est effectuée avec 3 oligonucléotides (a, b et c) situés de part et d'autres de sites HindIII. Le nombre de copies amplifiables est comparé avec celui obtenu pour un couple d'oligonucléotides ne chevauchant pas de site HindIII (ctrl), permettant ainsi de mesurer le taux de digestion. Les couples d'oligonucléotides a, b et c ciblent des régions de chromatine plutôt ouvertes et plutôt fermées. Le taux de digestion est ici d'environ 70%.

ne chevauchant pas de sites HindIII, permettant ainsi d'estimer le taux de coupure (figure 27B). Ce taux de coupure a été quantifié sur plusieurs sites HindIII pour s'assurer que la chromatine ouverte n'est pas significativement plus digérée que la chromatine fermée. Le couple d'oligonucléotides (a) est localisé dans un gène au niveau d'une région de chromatine acétylée, tandis que les couples (b) et (c) sont localisés dans une chromatine non-ouverte (pas d'acétylation, ni de H3K4me3), à respectivement 140 et 50 kb du gène le plus proche. Bien que correspondant à des contextes chromatinien différents, le taux de coupure de ces trois sites HindIII est similaire (70%). On peut donc raisonnablement conclure que le taux de digestion est d'environ 70% chez les cellules lymphoblastoïdes. Cette valeur est jugée suffisante pour poursuivre le protocole [Hagège et al., 2007].

c- Contrôle de la ligation et de la digestion par Hin1II

Après la digestion par HindIII, la ligation est une étape critique. Pour contrôler son efficacité, une aliquote (1.5 µL, après la ligation et la purification) est déposée sur gel (figure 28). Si la ligation s'est convenablement déroulée, l'ADN migre en une bande unique [Louwers et al., 2009] : c'est ce qui est obtenu. La digestion par Hin1II peut être également contrôlée par dépôt de 5 µL sur un gel : la majorité des fragments doivent être inférieurs à 1 kb, avec une densité optique maximale autour de 400 bp.

Ces contrôles ne sont pas quantitatifs mais, comparé à la digestion sur chromatine cross-linkée, ces étapes se déroulent généralement bien.

Le protocole 4C nécessite de nombreuses étapes enzymatiques, dont le déroulement peut être aisément suivi. Ces contrôles intermédiaires ont systématiquement été effectués afin de s'assurer de la bonne progression du protocole.

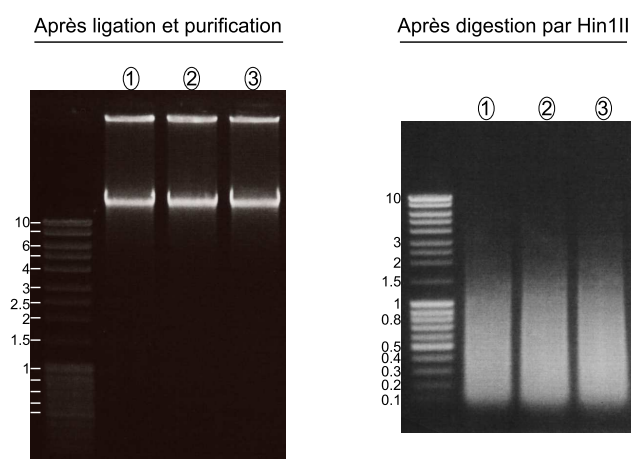


FIGURE 28 – Contrôle de la ligation et de la digestion par Hin1II.

3.2.2- Positionnement et conception des oligonucléotides 4C

- Plusieurs points de vue ont été choisis dans la région chr5:150-170Mb (figure 29A).
 - P1, P2, P3, P4 et P5 sont sur des pics de timing. Notons que P1, P2 et P3 sont très précoces, et que P4 et P5 sont plutôt répliqués en milieu de phase S (un peu avant pour P5, un peu après pour P4).
 - V1, V2, V3 et V4 sont dans des vallées de timing. Tous sont répliqués en fin de phase S sauf V2 qui est dupliqué en milieu de phase S.
 - S est situé entre une région précoce et une vallée similaire à V2.
 L'analyse des partenaires de ces différents points de vue 4C permet de tester l'ensemble des questions soulevées en fin du chapitre précédent.
- Ces points de vue sélectionnés, les oligonucléotides ont été placés en respectant deux règles. Premièrement, la distance entre le site HindIII et le site Hin1II le plus proche doit être la plus grande possible pour permettre la circularisation des molécules. Hin1II coupant en moyenne tous les 256 bp, nous avons choisi des fragments de restriction où la distance HindIII – Hin1II est supérieure à cette valeur (min = 281 bp pour P5, max = 1174 bp pour P4). Deuxièmement, le produit PCR qui est amplifié si le site HindIII n'est pas digéré doit avoir une taille compatible avec sa visualisation sur gel (contrôle qualité de l'amplification PCR). Pour les sites choisis, le produit de PCR "non-digéré" a une taille d'au moins 110 bp. Les tailles des fragments de restriction autour de V3 sont donnés en exemple figure 29B.
- La stratégie retenue pour coupler 4C et séquençage haut-débit est similaire à celle développée par [Splinter et al. \[2011\]](#) et [Noordermeer et al. \[2011b\]](#). Elle consiste à ajouter les adaptateurs requis pour le séquenceur Solexa en 5' des amorces 4C (figure 29C). Les premiers nucléotides lus lors du séquençage correspondront donc à la partie "spécifique" de l'amorce 4C-Seq à proximité du site HindIII. En conséquence, cette amorce a été placée au plus près du site en question.

3.2.3- Profils des PCR 4C

Protocole : PCR 4C

Les PCR sont réalisées à l'aide du kit Expand Long template PCR system de Roche avec le tampon #1, 200 μ M dNTP, 0.4 μ M amorces, 0.75U enzymes, 200 ng d'ADN. Le cycle PCR est [15s à 94°C, 1min à la température d'hybridation et 3min à 68°C]×36. 20 μ L de chaque PCR sont déposés sur un gel 1.8% et le gel est scanné grâce au Typhoon Scanner (532 nm).

Pour les PCR 4C-Seq, 10 réplicats sont effectués avec seulement 30 cycles de PCR et des amorces purifiées sur gel dénaturant (PAGE purification). Les PCR sont purifiées par chromatographie (Roche High Pure PCR Product Purification) et quantifiée par fluorimétrie (QuBit dsDNA HS). La librairie pour le séquençage est ajustée à 2.5 ng/ μ L (\approx 10 nM).

- La PCR 4C permet d'amplifier en une seule fois toutes les séquences ayant été capturées au niveau du site HindIII. Par conséquent, le profil de migration consiste en un "smear" qui s'étend du front de migration à quelques

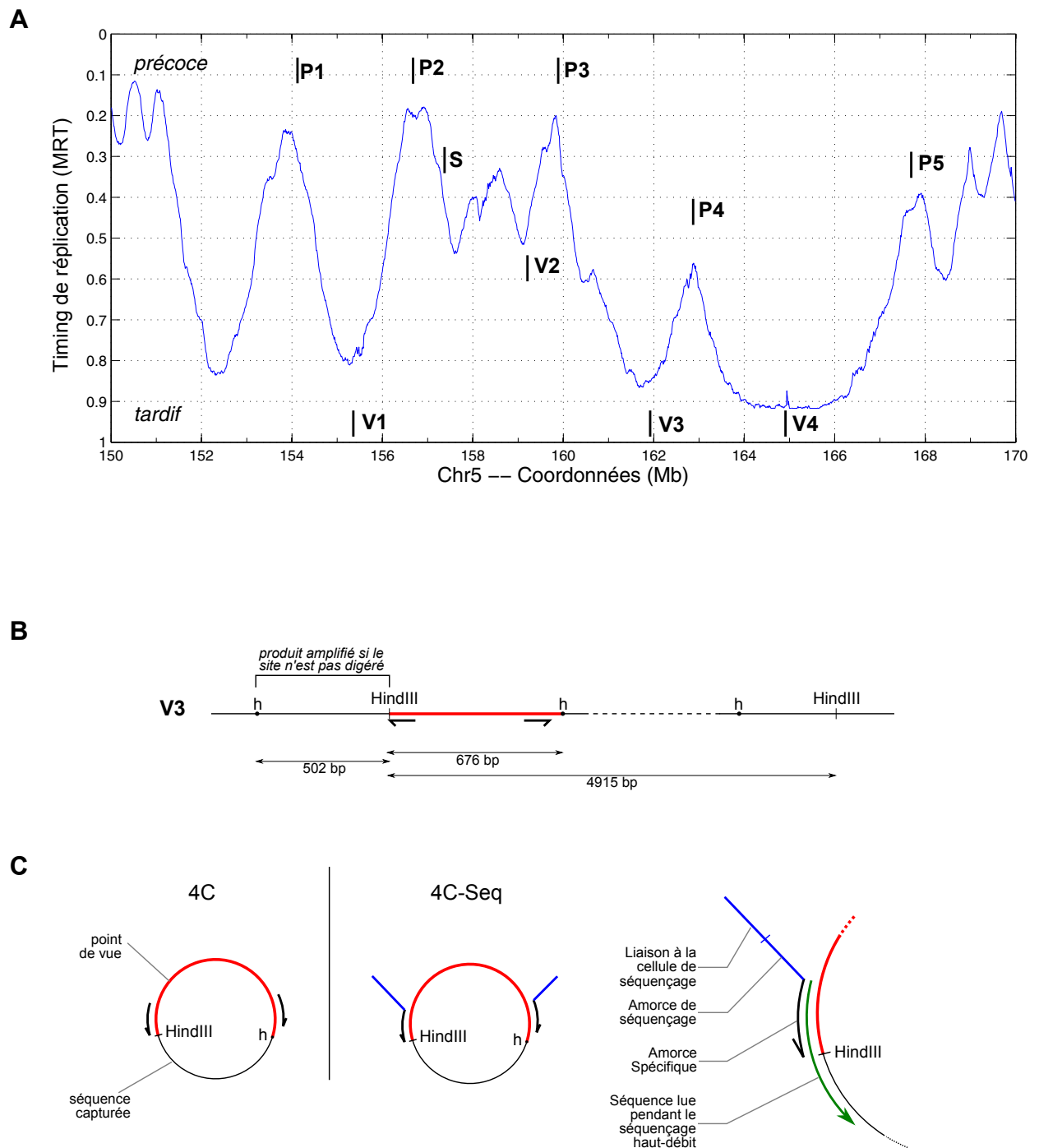


FIGURE 29 – Positionnement des amorces 4C. A- Locus sélectionné et points de vue choisis. Le timing représenté (MRT) correspond à celui établi par [Hansen et al. \[2010\]](#) et post-traité par [Baker et al. \[2012\]](#). B- Détail de la région V3 et taille des fragments de restriction. Le sigle "h" indique les sites *HindIII*. C- Amorces autorisant une analyse par séquençage haut-débit de type Solexa (4C-Seq). Les amorces sont divergentes (voir B) et permettent l'amplification de toute séquence capturée. Elles contiennent les adaptateurs requis pour le séquençage (bleu) et une partie spécifique du point de vue (noir). La séquence lue par le séquenceur est représentée en vert.

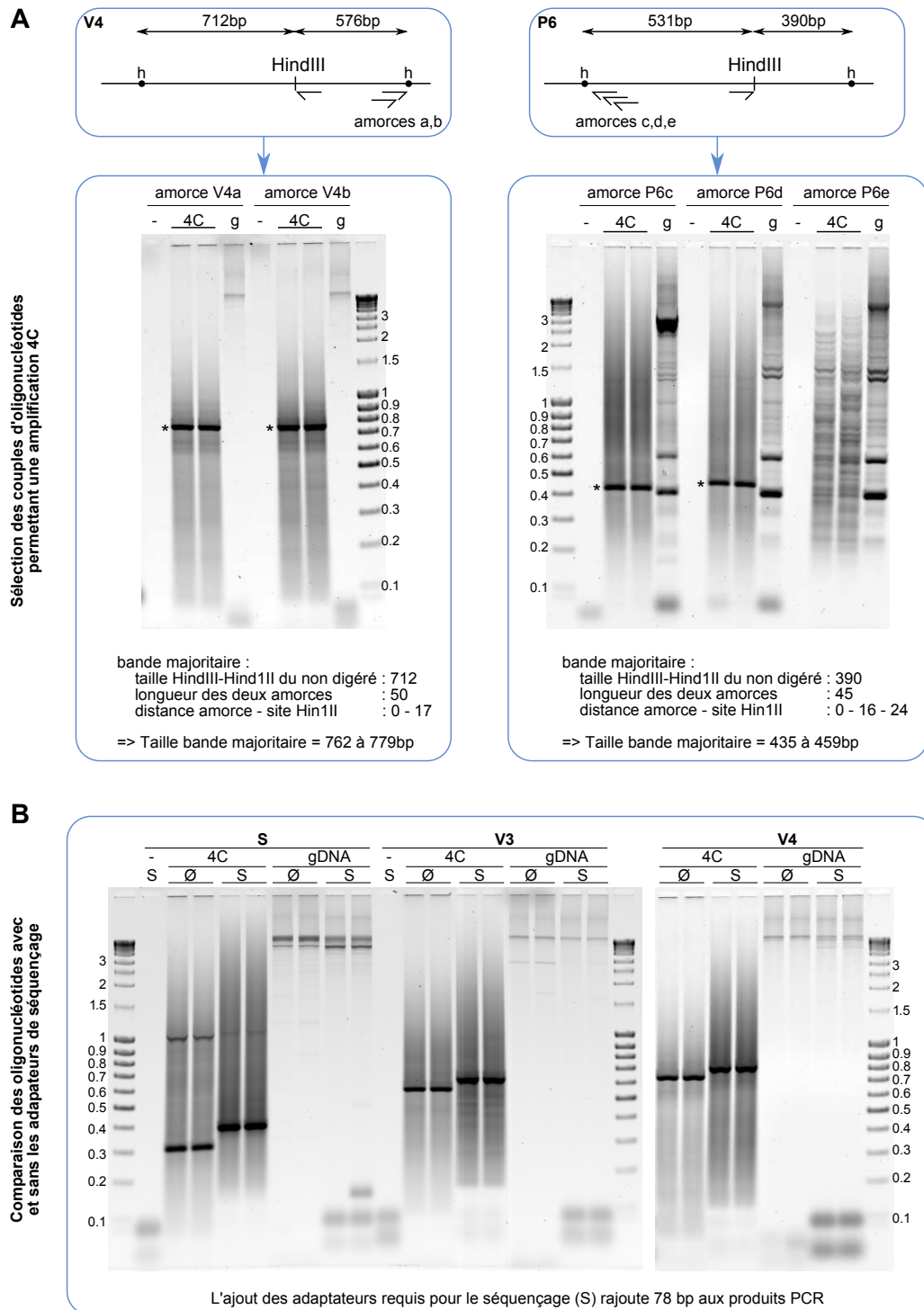


FIGURE 30 – Contrôle du bon déroulement des PCR 4C. A- Puisque la PCR 4C permet d'amplifier toutes les séquences capturées, le profil de migration est un "smear", avec une bande majoritaire (astérisque) correspondant au produit obtenu lorsque le site HindIII n'est pas digéré. A gauche, la PCR se déroule convenablement (région V4). A droite, le profil 4C peut être normal (amorces P6c, P6d) ou chaotique et non reproductible (amorce P6e), mais dans les 3 cas, les couples utilisés conduisent à une amplification sur de l'ADN génomique (g) ; ces couples ont donc été éliminés. B- L'ajout des adaptateurs de séquençage décale le "smear" et la bande majoritaire de 78 bp (taille des adaptateurs répartie sur les 2 amorces) mais ne change pas le patron globale de migration. Les couples comparés (sans adaptateur -Ø- ou avec adaptateur -S-) correspondent aux régions S, V3 et V4. En B, les amorces utilisées pour V4 correspondent à celles nommées V4a dans le panneau A.

kb (figure 30A, gauche). L'amplification PCR par certains couples (amorce P6e, figure 30A, droite) ne montre pas ce type de patron de migration : les amorces correspondantes ont donc été éliminées.

- Pour comprendre pourquoi certains couples d'amorces se révèlent inadaptés, il faut garder à l'esprit que l'amplification simultanée de nombreuses séquences soulève les mêmes problèmes techniques que les PCR multiplexes : des compétitions peuvent survenir, entraîner aléatoirement l'amplification préférentielle d'une cible plutôt qu'une autre [Markoulatos et al., 2002], et donc conduire à des amplifications stochastiques. Nous avons donc systématiquement testé la robustesse et la reproductibilité des amplifications 4C en effectuant des répliquats. Les amorces de la zone V4 (figure 30A gauche et 30B droite) donnent des profils extrêmement similaires :
 - immédiatement sous la bande majoritaire, on retrouve deux bandes fines entre 600 et 700 bp, visibles dans toutes les pistes ;
 - entre 500 et 600 bp, une zone plus claire est également observée.
 L'existence de profils de migration reproductibles indique que l'amplification est robuste.
- Les deux amorces 4C étant divergentes, aucune amplification ne devrait survenir sur de l'ADN génomique. Néanmoins, certains couples (amorces P6c/d/e figure 30A, droite) conduisent à des amplifications sur de l'ADN génomique. Ces couples ont été éliminés. Ainsi, malgré un profil 4C convenable les couples d'amorces P6c/d n'ont pas été utilisés par la suite (figure 30A, droite).
- Ajouter les adaptateurs requis pour le séquençage de type Solexa sur les amorces 4C revient à ajouter 54 nt en 5' de l'amorce placée sur le site HindIII et 24 nt en 5' de l'amorce à proximité du site Hin1II (soit 78 nt en tout). Pour s'assurer que ce doublement ou triplement de la taille des amorces n'affecte pas l'amplification PCR, les amplifications sans adaptateur (Ø) et avec adaptateur (S) ont été comparées (figure 30B). Les profils de migration montrent une forte ressemblance : la bande majoritaire est approximativement décalée de quelques bp (moins de 100 bp), tout comme l'ensemble du smear ; la zone claire entre 500 et 600 bp observée avec les amorces dans la zone V4 est elle aussi décalée ; il n'y a toujours pas d'amplification sur de l'ADN génomique. Cette stabilité du profil est un indice fort pour affirmer que les mêmes séquences sont amplifiées avec et sans les adaptateurs de séquençage.
- Seuls des couples d'amorces répondant à l'ensemble de ces critères de qualité ont été utilisés par la suite. La séquence des amorces retenues pour chaque point de vue est donnée table 3 (chapitre 4, page 144).

3.2.4- Des PCR 4C-Seq au séquençage haut-débit

a- Répliquats et profils de migration des PCR 4C-Seq

Les PCR 4C sont réalisées avec 200 ng d'ADN, soit 60 000 équivalent génomes haploïdes. Dans chaque tube PCR, le site HindIII du point de vue aura donc pu capturer au maximum 60 000 partenaires (si le taux de digestion est de 100%) et plus raisonnablement 40 000 partenaires (si le taux de digestion est de 70%). Or, le génome humain contient 25 fois plus de sites HindIII. Si les réarrangements

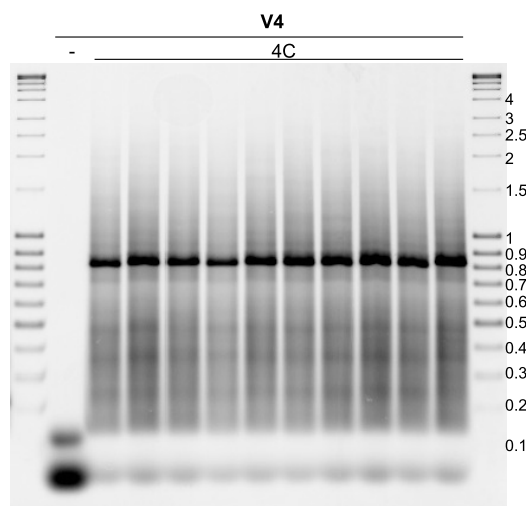


FIGURE 31 – 10 réplicats de la même PCR 4C-Seq. Ces 10 réplicats permettent de limiter les biais d'échantillonnage (voir texte). Les profils de migration de ces 10 réplicats sont extrêmement reproductibles, et correspondent au profil observé figure 30. A la différence de la figure 30, la PCR avant séquençage haut-débit ne contient que 30 cycles. L'amplification est donc moindre. Par conséquent, le profil de migration est observé par ajout de Sybr-Green dans le tampon de charge et le gel scanné à l'aide d'un laser bleu.

4C étaient aléatoires, seul un échantillon de toutes les recombinaisons possibles serait contenu dans les 200 ng d'ADN servant de matrice à la PCR. Pour limiter les biais d'échantillonnage, 10 réplicats pour chaque PCR 4C-Seq sont donc effectués (figure 31). Ainsi, au total 2 µg (soit 400 000 équivalents génomes haploïdes) sont utilisés comme matrice.

Pour éviter toute saturation, la PCR qui sera séquencée contient 6 cycles de moins que celles montrées figure 30. L'amplification étant moindre, la visualisation du profil de migration se fera avec du Sybr-green, plus sensible que le BET.

Les 10 réplicats montrent une très grande reproductibilité d'un tube à l'autre. Le gel de la figure 31 correspond aux couples d'amorces ciblant la région V4. En accord avec la figure 30, le smear s'étale de 150 bp à quelques kb et on retrouve une bande majoritaire avec en dessous une zone plus claire. Ces PCR étant reproductibles, elles sont rassemblées, purifiées par chromatographie et quantifiées par fluorescence.

b- Multiplexer les PCR 4C-Seq

Des tests préalables effectués dans l'équipe de Wouter de Laat ont montré que quelques millions de lectures par PCR 4C-Seq sont suffisantes pour couvrir la diversité des molécules contenue dans la PCR. En d'autres termes, le séquençage de 10 millions de molécules dans la PCR 4C-Seq n'identifiera pas plus de partenaires que le séquençage de 3 millions de molécules. Or, les séquenceurs haut-débit auxquels nous avons accès (Solexa Genome Analyzer IIx et Solexa HiSeq 2000) ont la capacité de séquencer plusieurs dizaines millions de molécules (30 millions pour le GAII, 80 millions pour le HiSeq 2000). Puisqu'un tel nombre de lectures n'est pas nécessaire en 4C, nous avons donc multiplexé les échantillons, en mélangeant 4 à 6 PCR 4C-Seq séquencées simultanément (nous verrons page 110 comment le séquençage est dé-multiplexé).

L'assertion selon laquelle quelques millions de lectures suffisent à couvrir la complexité des molécules des PCR 4C, peut être aisément vérifiée après-coup. En

3.2- De la librairie 4C au séquençage haut-débit

effet, nous pouvons comparer, dans nos 4C-Seq, le nombre de fragments HindIII différents capturés en fonction du nombre de lectures. Si quelques millions de lectures sont insuffisantes pour couvrir la diversité des librairies 4C, on s'attendrait à ce que la quantité de partenaires différents détectés augmente lorsqu'on accroît le nombre de lectures. Deux sets d'échantillons 4C séquencés avec des nombres de lectures différentes (l'un avec une moyenne de $2,6 \cdot 10^6$ lectures par PCR 4C et l'autre avec en moyenne $7,1 \cdot 10^6$ lectures par PCR 4C, soit un rapport de 2,7 entre les deux sets) ont donc été comparés. Le résultat montre que l'augmentation d'un facteur 2,7 du nombre de lectures n'est pas suivi d'une augmentation de la quantité de fragments de restriction différents capturés (le nombre de partenaires détectés reste en moyenne à 40 000). Nous pouvons conclure que $2,6 \cdot 10^6$ lectures suffisent à refléter la diversité des molécules présentes dans la PCR 4C-Seq.

Ce nombre de partenaires différents détectés (40 000) correspond exactement au nombre d'équivalents génomiques haploïdes présents dans un tube PCR (voir paragraphe précédent). N'effectuer qu'un seul réplicat de la PCR 4C-Seq (au lieu de 10, figure 31) aurait nécessairement conduit à un biais d'échantillonnage important.

L'analyse des PCR 4C-Seq ne nécessite pas un séquençage avec un grand nombre de lectures : nous avons donc multiplexé 4 à 6 PCR 4C avant le séquençage haut-débit.

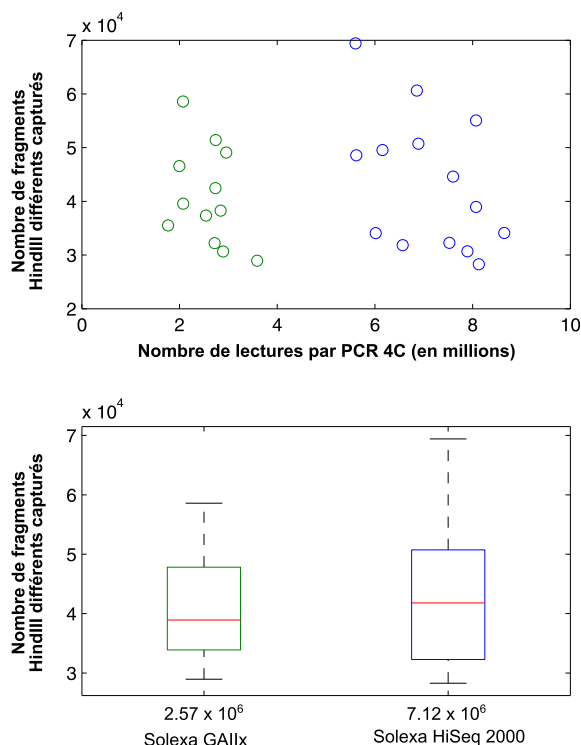


FIGURE 32 – Possibilité de multiplexer les PCR 4C avant séquençage. Multiplexer plusieurs PCR 4C revient à diminuer le nombre de lectures par PCR 4C. Pour contrôler que quelques millions de lectures sont suffisantes pour couvrir la complexité des PCR 4C, on compare le nombre de fragments HindIII différents capturés (soit le nombre de partenaires différents identifiés) avec le nombre de lectures. Les deux paramètres semblent relativement indépendants sur la gamme de lectures testée. En vert, les résultats obtenus avec le Solexa GAIIX ; en bleu, les résultats obtenus avec le Solexa HiSeq 2000.

c- Caractéristiques du séquençage haut-débit

Avec le design choisi des amorces 4C-Seq, les premiers nucléotides lus lors du séquençage haut débit correspondent à la partie "amorce spécifique" ciblant le point de vue (figure 29C). Ceci est utilisé pour démultiplexer le séquençage haut-débit : il suffit de trier les lectures en fonction de leur premiers nucléotides pour identifier le point de vue correspondant (voir page 110). Néanmoins, cela signifie que les premiers nucléotides ne sont pas informatifs pour déterminer le partenaire capturé. Les amorces 4C utilisées faisant en moyenne 25 nt de long, il faut donc séquencer au moins 55 bp afin de laisser 30 nt pour identifier le partenaire.

Le séquençage haut-débit effectué a donc une profondeur de 76 bp sur le Solexa GAIIX (Imagif, Gif-sur-Yvette) et de 100 bp sur le Solexa HiSeq2000 (MGX, Montpellier). Dans tous les cas, cela laisse suffisamment de nucléotides pour identifier la séquence capturée.

Les PCR 4C-Seq avec les adaptateurs requis pour le séquençage de type Solexa sont extrêmement reproductibles. 10 réplicats de chaque PCR ont été effectués pour minimiser les biais d'échantillonnage. 4 à 6 PCR 4C ont été mélangées avant le séquençage haut-débit qui a une profondeur de lecture d'au moins 76 bp.

3.3- Bioinformatique et traitement du séquençage haut-débit

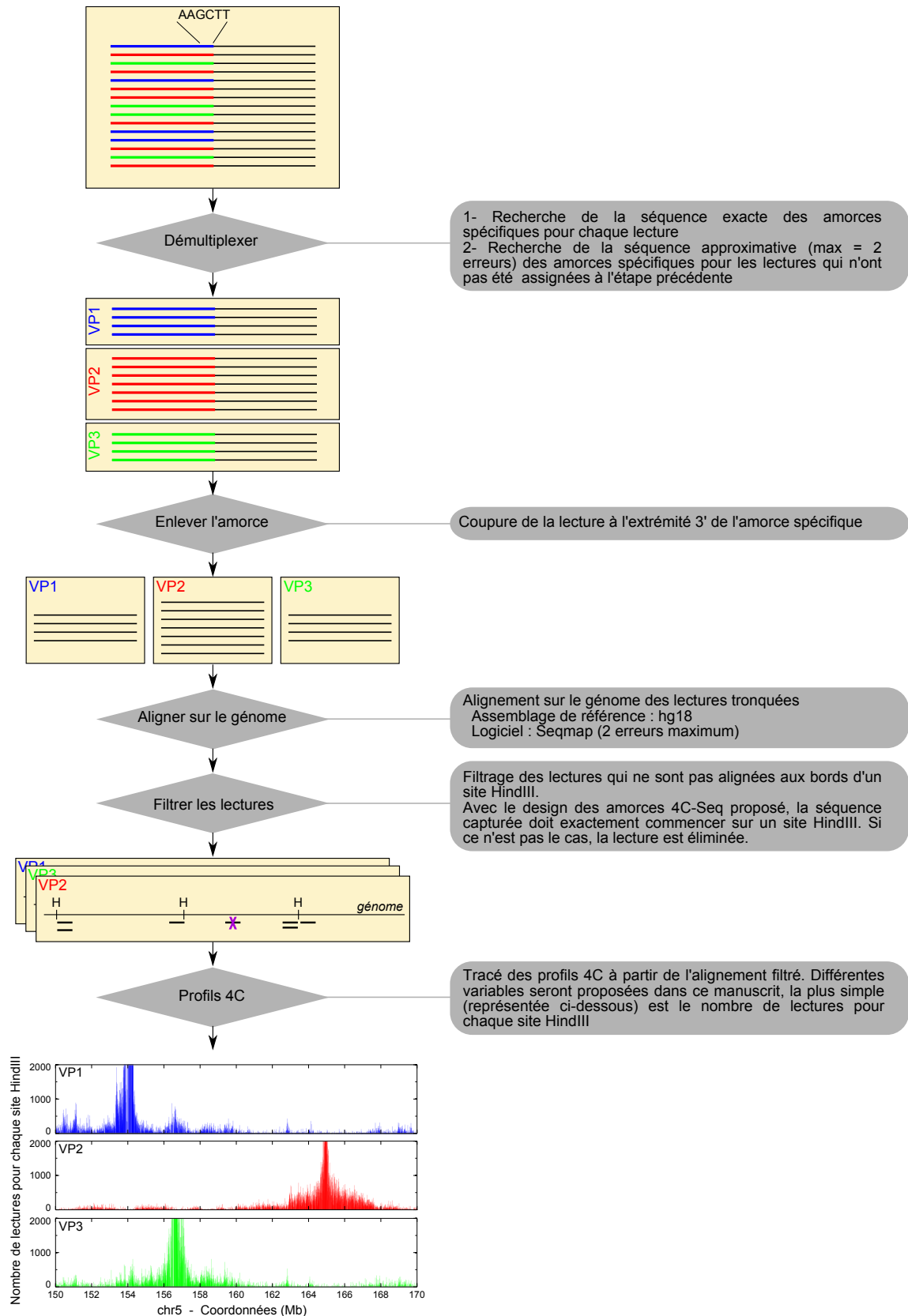
Les étapes requises pour passer de dizaines de millions de lectures résultant du séquençage haut-débit aux profils 4C sont résumées sur la figure 33. 5 étapes sont nécessaires.

1. Démultiplexer : la première étape consiste à démultiplexer les lectures. Cette étape permet de ré-attribuer à chaque point de vue les lectures correspondantes.
2. Enlever l'amorce : des séquences lues, on ne garde que la partie correspondant au partenaire capturé.
3. Aligner sur le génome : cette étape est réalisée par une application tierce (SeqMap [Jiang and Wong, 2008]).
4. Filtrer les lectures qui ne sont pas immédiatement flanquées d'un site HindIII.
5. Tracer les profils 4C

Dans cette section, nous allons détailler le principe et les résultats de chaque étape.

3.3.1- Démultiplexer les lectures 4C et conserver la partie spécifique du partenaire capturé

Comme mentionné ci-avant, les PCR 4C-Seq de 4 à 6 points de vue sont mélangées et séquencées simultanément. L'étape de démultiplexage permet de ré-attribuer à chaque point de vue les lectures correspondantes. Pour cela, on se base



sur les premiers nucléotides séquencés qui correspondent à la séquence de des amorces spécifiques (figure 29C). Le démultiplexage s'effectue en deux étapes : dans un premier temps, on trie les lectures par recherche de la séquence exacte des amorces spécifiques (alignement exact), puis, pour les lectures ne contenant pas la séquence exacte d'une amorce 4C, une recherche d'un alignement approximatif est effectuée en n'autorisant que deux erreurs (substitution, insertion et délétion).

En moyenne, la première étape permet d'attribuer 94% des lectures à leur point de vue respectif. Les 6% restantes contiennent donc des erreurs (introduites lors de l'amplification par PCR, lors du séquençage lui-même ou lors de la synthèse de l'amorce 4C) qui ne permettent pas l'identification de la séquence exacte du point de vue.

A l'issue de la seconde étape (recherche de séquences approximatives), seules 1,6% de toutes les lectures n'ont pas été associées à leur point de vue respectif car elles possèdent plus de 2 erreurs par rapport aux séquences des amorces spécifiques. Notons que ce pourcentage de lectures non attribuées est différent d'un séquenceur haut-débit à l'autre : sur le Solexa GAIIX (ImaGif, Gif-s-Yvette), seules 0,26% des lectures demeurent non-attribuées, alors que ce pourcentage est de 2,7% sur le Solexa HiSeq2000 (MGX, Montpellier). Il y a donc vraisemblablement plus d'erreurs introduites sur le Solexa HiSeq2000 que sur le GAIIX de Gif-s-Yvette.

Une fois le démultiplexage effectué, seule la partie 3' de la lecture doit être conservée pour être alignée sur le génome et permettre l'identification du partenaire capturé. La jonction entre la partie correspondant à l'amorce spécifique et la partie correspondant au partenaire capturé est déterminée lors du démultiplexage.

L'opération de démultiplexage permet d'attribuer la très grande majorité des lectures à leur point de vue correspondant. Seules 1,6% des lectures demeurent non-attribuées.

3.3.2- Alignement sur le génome

L'alignement sur le génome est effectué grâce au programme SeqMap [Jiang and Wong, 2008]. La séquence de référence utilisée pour le génome humain est l'assemblage hg18 de l'université de Californie, Santa Cruz (UCSC, genome.ucsc.edu). Pour les cellules lymphoblastoïdes et les PBMC (Cellules mononucléées du sang périphérique), en moyenne 88,8% des lectures peuvent être alignées sur le génome humain et une grande majorité sont retrouvées à l'identique (figure 34A). Ceci témoigne de la robustesse du séquençage, mais aussi de la fidélité de l'amplification par PCR. Pour ~6% des lectures, le logiciel ne parvient pas à identifier de localisation. Ces lectures contiennent vraisemblablement trop d'erreurs pour être alignées ou correspondent à des régions non séquencées. Enfin, un peu plus de 5% des lectures sont retrouvées à de multiples positions dans le génome humain : elles correspondent en partie aux partenaires localisés dans des éléments répétés. Ces résultats correspondent à une moyenne des trois types cellulaires utilisés, néanmoins chacun présente une distribution analogue. En conclusion, quelque soit le type cellulaire considéré, près de 89% des lectures sont exploitables.

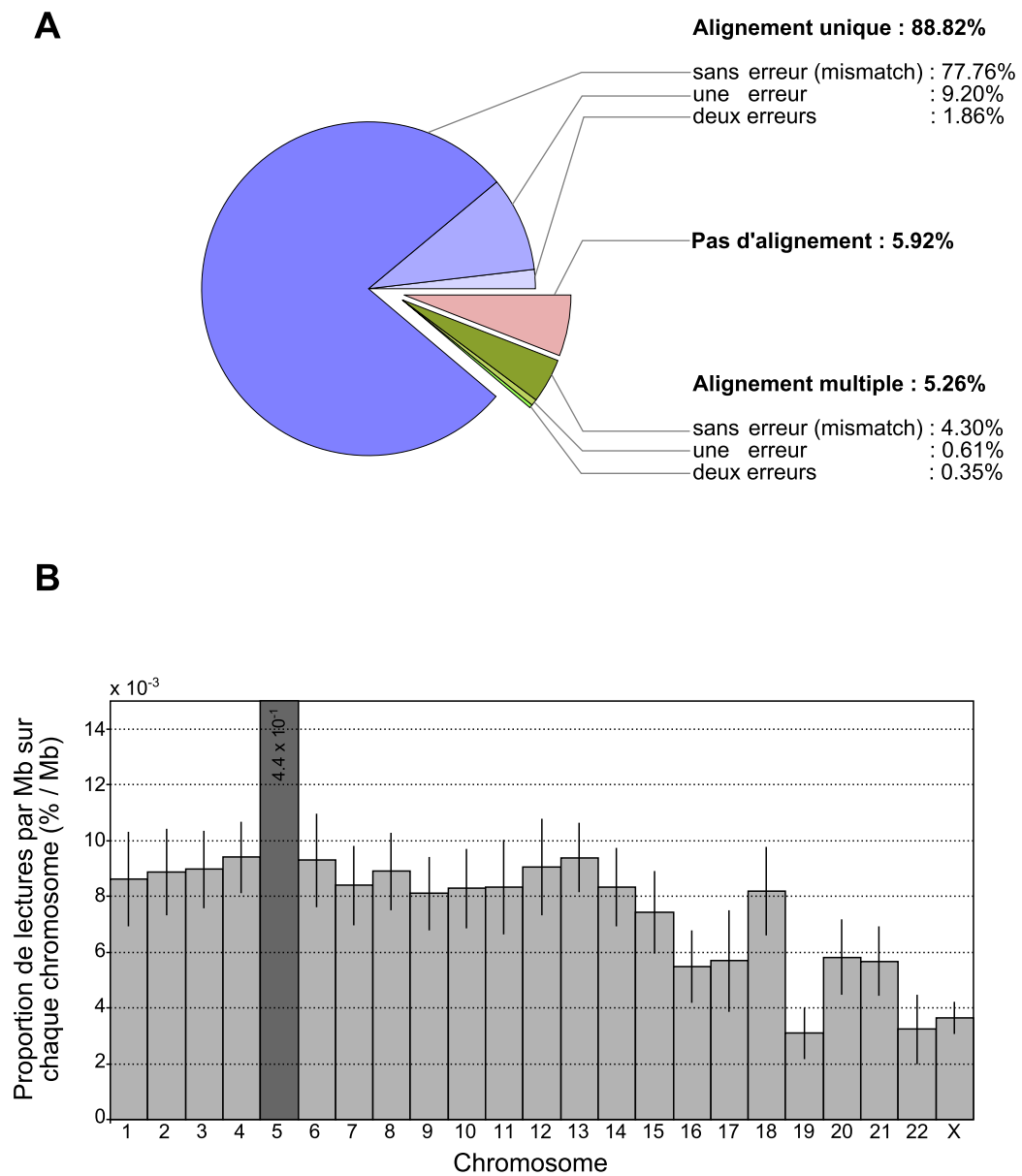


FIGURE 34 – Alignement sur le génome humain.

A- Plus de 88,8% des lectures des 4C ont été alignées sur le génome humain de référence (hg18). A peu près 6% des lectures ne peuvent être alignées, et 5,2% sont retrouvées à de multiples positions dans le génome. Ces valeurs correspondent à une moyenne pour tous les points de vue pour les cellules lymphoblastoïdes et les PBMC.

B- La proportion de lectures par Mb sur chaque chromosome est représentée (pourcentage sur chaque chromosome divisé par la taille séquencée de chaque chromosome). Ainsi, 79% des lectures sont localisées sur le chromosome 5 (les interactions intra-chromosomiques prédominent) qui a une taille séquencée de 177 Mb, soit 0.44% de toutes les lectures par mégabase. Les 21% restantes se répartissent sur les autres chromosomes. Les données sont exprimées en moyenne \pm écart-type pour les 10 points de vue et les 3 types cellulaires.

Avant d'étudier précisément le positionnement des lectures alignées, on peut analyser de quel chromosome elles proviennent. Ainsi, on observe que $79 \pm 3,3\%$ des lectures (moyenne \pm écart-type) se localisent sur le chromosome 5 (les lectures correspondant aux cas où le site HindIII n'est pas digéré –bande majoritaire, voir figure 30– ont au préalable été filtrées pour obtenir ce pourcentage). Ce résultat démontre que la majorité des contacts chromatinien s'effectuent au sein du chromosome 5. Ce résultat est en accord avec l'existence de territoires chromosomiques occupant un emplacement discret du noyau et relativement isolés les uns des autres. Les 21% des lectures restant se répartissent entre les autres chromosomes.

Afin d'étudier la répartition des lectures, nous avons quantifié leur densité par mégabase. Par exemple, 1,94% des lectures sont localisées sur le chromosome 1 qui fait 225 Mb de long (taille séquencée). Ceci signifie qu'on retrouve $\frac{1,94}{225} = 8,6.10^{-4}\%$ de toutes les lectures par mégabase de chromosome 1 (figure 34B). Cette mesure de la proportion de lectures normalisée par la taille du chromosome permet de comparer la densité en lectures pour différents chromosomes. Cette densité est plus forte pour les grands chromosomes (chr 1 à chr 14) et plus faible pour les petits (chr 15 à chr 22). En moyenne, les points de vue choisis sur le chromosome 5 interagissent donc plus avec des partenaires des grands chromosomes qu'avec des partenaires des petits chromosomes. Seul le chr 18 échappe à cette règle : la densité en lecture y est proche de celle observée sur les grands chromosomes. Ces différentes observations peuvent être expliquées par l'agencement radial non-aléatoire des territoires chromosomiques (voir page 35) qui conduit à une localisation en périphérie du noyau des grands chromosomes pauvres en gènes (et du chr 18) et à une localisation au centre du noyau des petits chromosomes denses en gènes. Le chr 5, grand chromosome, est donc plutôt retrouvé en périphérie du noyau, là où sont situés les autres grands chromosomes. Il est donc cohérent d'observer que les points de vue choisis du chr 5 interagissent plus avec d'autres grands chromosomes. En plus de confirmer l'agencement non radial des territoires chromosomiques, cette observation est un indice fort de la qualité du 4C-Seq.

Le chromosome X a une taille proche de celle du chromosome 7, pourtant la densité en lectures sur ce chromosome est deux fois plus basse que celle observée sur les grands chromosomes (figure 34B). Ceci s'explique simplement par le caryotypes des cellules utilisées dans cette étude. Les cellules lymphoblastoïdes, PBMC dormants et PBMC cyclants sont tous issus d'individus masculins, et n'ont donc qu'une seule copie du chromosome X. La probabilité que le chr 5 interagisse avec le chr X est donc deux fois moindre, ce qui explique que la densité en lectures est deux fois plus faible.

En accord avec l'existence de territoires chromosomiques et leur arrangement radial dans le noyau, 79% des contacts s'effectuent au sein du chr 5, et les 21% restant sont établis de préférence avec d'autres grands chromosomes. Ce résultat atteste de la qualité du 4C-Seq.

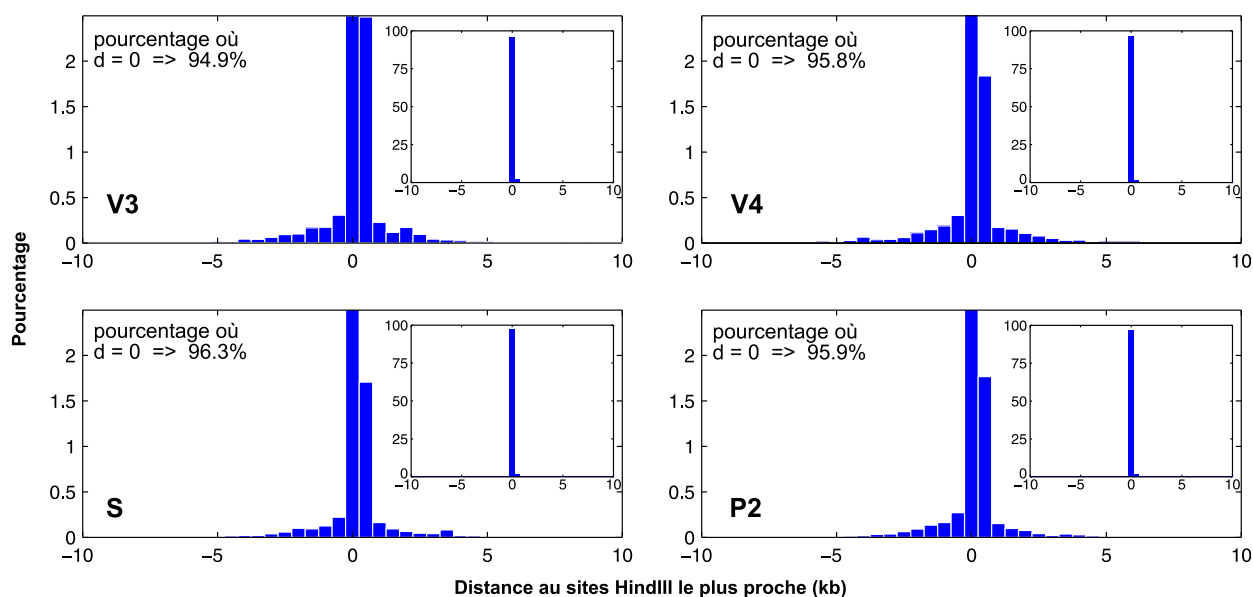


FIGURE 35 – Alignement sur le génome - distance au site HindIII le plus proche. La distance entre les lectures alignées sur le génome humain et le site HindIII le plus proche a été mesurée. Si le 4C-Seq se déroule convenablement, la lecture alignée doit être immédiatement adjacente à un site HindIII. Les distances mesurées sont représentées sous forme d'histogrammes exprimés en pourcentage. Les encarts correspondent à la pleine dynamique de l'axe des ordonnées ; pour les panneaux principaux, l'axe des ordonnées a été redimensionné (ordonnée maximale = 2,5%).

3.3.3- Filtrage des lectures alignées sur le génome en fonction de leur distance au site HindIII le plus proche

Dans l'expérimentation 4C, les recombinaisons s'effectuent au niveau d'un site de restriction (dans notre cas, HindIII). Les partenaires identifiés par alignement sur le génome humain doivent alors être situés à proximité d'un site HindIII. Pour le vérifier, la distance entre le partenaire et le site HindIII le plus proche a été mesurée.

Cette analyse montre que la plupart des partenaires capturés sont situés à moins de 250 bp du site HindIII le plus proche (figure 35 pour les points de vue V3, V4, S et P2 dans les cellules lymphoblastoïdes ; les autres points de vue et les autres cellules montrent des distributions analogues). Conformément à ce qui est attendu, le partenaire se situe même, dans la grande majorité des cas, directement sur un site HindIII. En effet, $96,8 \pm 0,7$ % des séquences alignées (moyenne \pm écart-type) sont directement situées sur un site HindIII ($d = 0$, figure 35). Ce pourcentage très élevé est un indice fort pour le bon déroulement du 4C (préparation des bibliothèques, PCR et séquençage haut-débit).

La position attendue pour les partenaires identifiés étant juste à côté d'un site HindIII, toutes les lectures qui ne sont pas directement alignées à cette position (distance différente de zéro) ont été écartées des analyses ultérieures.

Conformément à ce qui est attendu, la plupart des lectures alignées sur le génome humain se localisent directement sur un site HindIII. Toutes lectures retrouvées en dehors d'un site HindIII ont été écartées des analyses ultérieures.

Conclusion

Ce chapitre nous a permis de détailler chaque étape du protocole 4C-Seq, depuis la préparation des librairies 4C jusqu'à l'alignement des lectures sur le génome de référence. L'une des étapes les plus critiques dans la préparation des librairies 4C consiste à atteindre un taux de digestion élevé de la chromatine cross-linkée (comme ici de 70%) afin d'autoriser une capture efficace des contacts chromatinien. Ce taux de digestion peut aisément être contrôlé par migration sur gel et par PCR quantitative.

L'exploitation des librairies 4C nécessite l'utilisation d'amorces suffisamment robustes pour éviter toute amplification stochastique. Si la conception des amorces 4C répond aux mêmes critères que celle de n'importe quelle amorce de PCR, la sélection des amorces réellement robustes sera majoritairement empirique : la migration des PCR 4C sur gel doit être reproductible d'un réplicat à l'autre et contenir une bande majoritaire correspondant au cas où le site HindIII ciblé n'est pas digéré. Contrôler qu'aucune amplification ne survient si le génome n'a pas été circularisé est également essentiel.

Puisque quelques millions de lectures sont suffisantes pour couvrir la complexité des PCR 4C-Seq, plusieurs PCR ont été multiplexées et séquencées simultanément. Les algorithmes de démultiplexage permettent l'attribution de chaque lecture dans au moins 97% des cas, en se basant sur la reconnaissance du point de vue par les premiers nucléotides de chaque lecture. Ce tri effectué, les lectures sont alignées sur le génome de référence. L'alignement montre que, conformément à ce qui est attendu, les lectures sont localisées sur des sites HindIII dans 95% des cas. Leur répartition par chromosome montre enfin que les contacts chromatinien s'effectuent en accord avec l'existence de territoires chromosomiques, ainsi qu'avec leur agencement radial dans le noyau.

Ces 4C-Seq ont été effectués à partir de cellules lymphoblastoïdes et de cellules mononucléées du sang périphérique. Les profils 4C correspondants à ces types cellulaires seront présentés dans le chapitre 4. Dans les deux dernières semaines avant impression de ce manuscrit, nous avons reçu le résultat d'un 4C-Seq effectué dans les cellules K562 (cellules issues d'une leucémie myéloïde à caractère érythrocytaire). Le même protocole expérimental a été suivi, avec les mêmes contrôles, et la même bioinformatique. Les profils correspondants seront présentés dans le chapitre 5.

Correspondance entre timing de réplication et organisation de la chromatine

Problématique :

Il est communément admis que l'organisation de la chromatine influe sur le contrôle de l'expression génique. Cet effet peut impliquer tous les niveaux de repliement hiérarchique de la chromatine. En effet, certaines modifications post-traductionnelles des histones sont plutôt retrouvées au niveau de gènes exprimés et d'autres au niveau de gènes réprimés. De plus, la chromatine est, dans l'ensemble, plus compactée dans les régions inactives. Enfin, l'établissement de certains contacts à longue-distance (> 100 kb) semble essentiel pour le contrôle fin de l'expression génique (chapitre 1).

Si l'implication de ces différents niveaux de repliement sur la transcription est bien établi, leurs relations avec la réplication de l'ADN sont nettement plus flous. Certes, la progression des fourches de réplication pourrait façonner l'agencement de la chromatine par l'ancrage de certaines séquences au niveau d'usines de réplication, mais les preuves directes d'une interaction entre séquences répliquées de manière synchrone manquent (chapitre 2).

Pour étudier les liens entre agencement de la chromatine et réplication, nous avons employé une méthodologie de type 4C-Seq. Celle-ci nous a permis d'identifier avec précision les partenaires chromatiniens de séquences répliquées en début ou en fin de phase S. Les nombreuses étapes du protocole ont été abordées dans le chapitre 3 et le bon déroulement de chacune a été finement contrôlé. En sortie, nous disposons de plusieurs millions de lectures représentant les partenaires de 10 points de vue dans 3 types cellulaires (cellules lymphoblastoïdes, Peripheral Blood Mononuclear Cell –PBMC– dormantes, et PBMC cyclantes). Dans ce chapitre-ci, je vous propose d'analyser comment se distribuent, le long du génome, les partenaires de loci répliqués en début ou en fin de phase S.

Résumé des résultats :

► *Un U-domaine de timing, définit par ses deux zones d'initiation précoces encadrant une vallée de timing, forme un domaine structural du noyau.* Ceci est observé grâce aux points de vue 4C situés au centre d'une vallée de timing : leur capacité d'interaction chute subitement lorsque le partenaire contacté est situé au-delà d'un pic de timing.

► *Les zones d'initiation précoce interagissent préférentiellement entre-elles.* Ceci est observé par 4C, aussi bien entre deux zones d'initiation précoce consécutives

(séparation génomique de l'ordre du mégabase), qu'entre régions précoces distantes de plusieurs dizaines, voire centaines de mégabases. Les profils 4C des régions précoces suivent admirablement le timing de réplication le long du chromosome, ce qui n'avait encore jamais été observé.

➤ *Les zones répliquées en fin de phase S interagissent avec leurs homologues.* Combiné au résultat précédent, nous pouvons donc conclure que loci précoces et tardifs occupent des territoires nucléaires distincts.

➤ *Les loci précoces sont plus proches les uns des autres qu'attendu.* Les résultats 4C sont corroborés par des mesures de distances par hybridation *in-situ* qui montrent que les loci précoces sont plus proches dans le volume nucléaire que l'on pourrait s'y attendre si le repliement était aléatoire. La chromatine est donc organisée de telle sorte à rapprocher les séquences répliquées en début de phase S.

➤ *Cette organisation existe en phase G0.* Les profils 4C de cellules lymphoblastoïdes, de PBMC en cycle et de PBMC dormantes (bloquées en phase G0) sont extrêmement similaires. L'organisation particulière de la chromatine autour de la réplication n'existe donc pas uniquement en phase S : elle persiste vraisemblablement tout au long du cycle cellulaire.

Soumission de ce travail :

Ce travail a été soumis à Genome Research en Février 2012. Vous trouverez ci-après la publication telle qu'elle a été soumise (à une ou deux références à des figures près). Afin de simplifier la lecture, l'article a été mis en page de telle sorte que la figure et le texte s'y rapportant soient, dans la mesure du possible, l'un en face de l'autre. Les figures/tableaux supplémentaires ont été inclus de la même façon. Pour les distinguer des figures principales, le titre des figures/tableaux supplémentaires commencent par (*suppl.*)¹.

1. Les éditeurs de Genome Research n'ayant pas accepté ce manuscrit (après évaluation par les pairs), une version légèrement modifiée de cet article a été proposée et acceptée par Nucleic Acids Research.

DNA replication timing delineates structural domains in the nucleus of cycling and resting human cells

Benoit Moindrot¹, Benjamin Audit¹, Petra Klous³, Antoine Baker¹, Claude Thermes², Wouter de Laat³, Philippe Bouvet¹, Fabien Mongelard¹ and Alain Arneodo^{1*}

¹ Laboratoire Joliot-Curie, Ecole Normale Supérieure de Lyon, CNRS, F-69007 Lyon, France ; Université de Lyon, F-69000 Lyon, France

² Centre de Génétique Moléculaire (CNRS UPR3404), Allée de la Terrasse, 91198 Gif-sur-Yvette, France

³ Hubrecht Institute-KNAW & University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

* Corresponding author : Email : alain.arneodo@ens-lyon.fr

Abstract:

Although chromatin folding is known to be of functional importance to control the gene expression program, little is known regarding its interplay with DNA replication. Here, using Circular Chromatin Conformation Capture (4C) combined with high-throughput sequencing, we study long-range chromatin interactions in relation with replication timing data. We demonstrate that early-initiation zones contact each other by looping out the intervening late-replicating loci. These early-initiation zones moreover coincide with frontiers limiting the interactions of neighboring late-replicated regions, providing direct evidence for replication-related boundaries of megabase-sized structural chromatin domains. We further show that, at larger distances, early- and late-replicated loci preferentially interact with their counterparts, exhibiting a striking correlation between the contact frequencies and the replication timing profile. This demonstrates that loci with opposite replication timing segregate in the nuclear space. The preferential interaction between neighboring early-replicated loci is also demonstrated by *in-situ* distance measurements showing their physical proximity. Highly similar results are obtained in a human lymphoblastoid cell line, as well as in cycling and resting peripheral blood mononuclear cells. We therefore provide direct clues for a robust chromatin organization, common to cycling and resting cells, that is tightly linked to DNA replication.

4.1- Introduction

The timely firing of replication origins is essential for most eukaryotic cells to allow the proper duplication of the genetic material during the S-phase. Replication origins are activated according to a specific temporal program that has been studied in various species allowing to distinguish sequences replicated in the early S-phase from sequences replicated later [Raghuraman et al., 2001; Schübeler et al., 2002; White et al., 2004; Woodfine et al., 2004; Huvet et al., 2007; Hiratani et al., 2008; Hansen et al., 2010; Yaffe et al., 2010; Chen et al., 2010]. Strikingly, the replication timing is globally stable over different cell lines, even if some loci may exhibit cell type-related timing differences; furthermore this stable timing pattern is also conserved, at least to some extent, between homologous regions of different species [Hiratani et al., 2008, 2010; Ryba et al., 2010; Yaffe et al., 2010; Chen et al., 2011]. The replication timing is significantly correlated with genomic and epigenetic features (reviewed in Farkash-Amar and Simon [2010]). For instance, most of the early-replicated regions are gene-rich, transcriptionally active and localized in open chromatin, while the late-replicated regions are mostly found in silent, gene-poor and closed chromatin [Schübeler et al., 2002; White et al., 2004; Woodfine et al., 2004; MacAlpine et al., 2004; Thurman et al., 2007; Hiratani et al., 2008; Audit et al., 2009; Ryba et al., 2010; Chen et al., 2010]. The replication timing thus reflects several genomic, epigenetic and functional characteristics.

Besides linear features of chromatin, DNA replication has been related to the 3D-organization of the nucleus. From the observation that thousands of replication forks are only found in hundreds of discrete BrdU-labeled foci, it has been proposed that DNA sequences replicating at the same time could gather [Ma et al., 1998; Jackson and Pombo, 1998; Cook, 1999; Berezney, 2002; Koberna et al., 2005], possibly *via* the anchoring of giant chromatin loops on the nuclear skeleton (reviewed in Anachkova et al. [2005], Ottaviani et al. [2008] and Arneodo et al. [2011]). Courbet et al. [2008] carried out elegant experiments showing that the size of the nuclear skeleton-anchored loops clearly depends on the velocity of the fork and on the density of initiations. They observed that experimental slowing down of replication forks is compensated for by the mobilization of latent replication origins in the subsequent S-phase, and by the formation of smaller chromatin loops. This argues in favor of a replication-related high-order folding of chromatin. The recent availability of replication timing data in several cell lines now offers the opportunity to test this idea more systematically and to address to which extent a structural counterpart exists to the functional replication domains identified from the replication timing. 3C methodologies [Dekker et al., 2002] have emerged as powerful strategies to analyze in detail the long-range folding of chromatin and by extension, the chromosome organization in the nucleus. Different genome-wide mapping strategies have been developed [de Wit and de Laat, 2012], including circular chromosome conformation capture (4C) for the highly detailed analysis of DNA contacts made by selected loci [Simonis et al., 2006; Zhao et al., 2006], and Hi-C for a more global pairwise comparison of contacts between all genomic loci [Lieberman-Aiden et al., 2009].

Hi-C data have already been compared to the replication timing of related cell lines [Ryba et al., 2010; Yaffe et al., 2010]. Ryba et al. [2010] observed a tight link between the replication timing and the Hi-C correlation matrices (principal eigenvector), showing that sequences localized near each other replicate at similar time. Moreover, Yaffe and coworkers numbered more interchromosomal interactions than expected between regions having similar replication timing [Yaffe

et al., 2010]. However, the Hi-C data do not always allow the direct visualization of the chromatin contacts for chosen loci, principally because of insufficient sequencing depth to cover all the combinatorial possibilities (figure 24 in chapter 2 and corresponding text). Several essential questions thus remain to be answered, such as the existence of preferential interacting partners, and whether they display a particular distribution along the genome that may reveal some specific, replication associated, chromatin organization. At what scales can we detect such an organization? Does this organization persist across the cell cycle? Does the existence of timing-dependent structural domains imply the existence of local boundaries restricting the range of 3D-interactions?

Here, we used the 4C methodology [Simonis et al., 2006] in order to identify with high precision the main partners of multiple early- and late-replicated regions and to determine to which extent the 3D-organization of chromatin can be linked to the replication timing. For this study, we focused on a large region (20 Mb) of the human genome in which the replication timing presents large fluctuations that are conserved in most examined cell types (figure 25A in chapter 3). In particular, these fluctuations define loci replicating earlier than their immediate surroundings (timing peaks), and conversely regions replicating later than the surroundings (timing valleys). 4C experiments were performed in a lymphoblastoid cell line, in resting and in cycling peripheral blood mononuclear cells (PBMC) using 10 viewpoints in the selected region. We observed, and confirmed by FISH experiments, that early timing peaks coincide with frontiers limiting the spreading of strong local interactions and delineating megabase-sized 3D-structural domains. In addition, at larger distances, profiles of long-range interactions from timing peaks viewpoints faithfully reproduce the replication timing profile, which sheds a new light on the propensity of early- and late-replicating regions to contact other chromosome regions. These results are robustly observed in cycling and G0-blocked cells, indicating that the observed organization exists in G0 and does not result from the ongoing DNA replication.

4.2- Results

We employed 4C in combination with high-throughput sequencing (4C-seq) [Splinter et al., 2011] to explore the correlation between replication timing and chromosome 3D-structure. 4C is used to capture, in the whole genome, the loci interacting with a chosen sequence, namely the viewpoint. We focused on a large (20 Mb) region of human chromosome 5 in which the replication timing profile measured in lymphoblastoid cell line presents large fluctuations and is highly conserved in different cell types (figure 25A in chapter 3). This allowed us to identify local timing peaks as regions that replicate earlier than their immediate surroundings, and conversely local timing valleys as regions replicating later than their surroundings. 4C viewpoints were then unambiguously positioned on some local timing peaks or in timing valleys (figure 37A). Five viewpoints were selected on early peaks (P1 to P5), 3 in very late valleys (V1, V3, V4), one in a valley of intermediate timing (mid S-phase, V2) and one in a transition from very early to mid S-phase (S). The local chromatin and genomic features around the viewpoints are shown in figure 36. A mean of 2.6 millions reads per viewpoint was generated in lymphoblastoid cells (table 1). About 76% of the mapped reads

Viewpoints	P1	P2	P3	P4	P5	S	V1	V2	V3	V4	(total)
Reads number (absolute)	2 740 247	2 719 730	2 070 107	2 733 911	3 589 694	2 841 857	2 955 540	2 891 073	2 075 915	1 766 323	26 384 397
Lymphoblastoid	2 541 407	8 127 305	1 994 584	10 781 467	14 167 304	7 523 709	8 069 640	12 079 074	6 968 336	7 893 763	79 746 589
Resting Lympho	6 890 248	6 015 350	5 603 514	8 071 340	11 488 271	7 601 659	6 857 206	8 646 373	5 617 122	6 153 403	72 944 486

Interval (in Mb) around the viewpoint containing 75% of chromosome 5 reads											(mean)
Lymphoblastoid	4.91	2.83	3.13	5.65	3.93	2.43	3.58	2.67	2.65	2.5	3.428
PHA Lympho	5.16	0.945	4.11	6.02	2.4	1.67	2.12	2.78	2.78	2.39	2.9775
Resting Lympho	7.72	1.56	6.57	11.01	7.14	5.09	4.53	4.54	6.17	5.9	6.023

Table 1: (suppl.) Read number and distribution. 4C-Seq reads were first processed to identify the viewpoints and then mapped on the human genome (UCSC hg18 assembly). This table summarizes the number of reads identified for every cell line and viewpoint. In addition, we provide the size of the interval centered on the viewpoint containing 75% of the whole chromosome 5 reads. This interval is a few Mb, confirming that most of the interactions captured are local interactions. Reads corresponding to non-digested product were not included for the calculation of the 75%-interval.

Viewpoints	chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	chr13	chr14	chr15	chr16	chr17	chr18	chr19	chr20	chr21	chr22	chrX
Resting PBMC																							
V1	1.56%	1.74%	1.65%	2.07%	80.46%	1.42%	1.17%	1.20%	0.96%	0.95%	0.96%	1.18%	1.00%	0.66%	0.44%	0.36%	0.33%	0.67%	0.13%	0.29%	0.19%	0.07%	0.55%
V2	1.63%	1.97%	1.76%	1.90%	79.40%	1.36%	1.28%	1.25%	0.99%	1.06%	1.04%	1.28%	0.93%	0.75%	0.59%	0.44%	0.39%	0.68%	0.12%	0.37%	0.21%	0.08%	0.50%
V3	1.69%	1.99%	1.72%	2.04%	79.01%	1.45%	1.35%	1.41%	1.02%	1.01%	1.13%	1.24%	1.02%	0.72%	0.45%	0.40%	0.29%	0.74%	0.15%	0.27%	0.20%	0.07%	0.61%
V4	1.70%	1.98%	1.56%	1.92%	79.75%	1.44%	1.30%	1.32%	1.01%	1.05%	1.06%	1.22%	0.94%	0.63%	0.45%	0.39%	0.28%	0.75%	0.12%	0.30%	0.19%	0.08%	0.55%
S	1.74%	2.02%	1.76%	1.92%	78.33%	1.58%	1.34%	1.34%	1.06%	1.02%	1.14%	1.24%	1.02%	0.73%	0.61%	0.49%	0.48%	0.63%	0.14%	0.38%	0.24%	0.10%	0.60%
P5	1.79%	2.05%	1.65%	1.70%	78.81%	1.34%	1.20%	1.19%	1.11%	1.20%	1.21%	1.38%	0.96%	0.74%	0.66%	0.63%	0.45%	0.77%	0.15%	0.44%	0.18%	0.08%	0.49%
P4	2.53%	2.65%	2.13%	1.99%	73.71%	1.86%	1.59%	1.65%	1.29%	1.53%	1.39%	1.65%	0.93%	0.87%	0.71%	0.58%	0.53%	0.76%	0.23%	0.47%	0.27%	0.09%	0.63%
P3	1.67%	1.83%	1.55%	1.55%	81.05%	1.25%	1.16%	1.21%	0.95%	1.05%	1.06%	1.18%	0.85%	0.62%	0.53%	0.41%	0.32%	0.59%	0.14%	0.30%	0.19%	0.06%	0.47%
P2	1.93%	1.91%	1.59%	1.47%	79.93%	1.32%	1.12%	1.17%	0.98%	1.11%	0.95%	1.27%	0.82%	0.75%	0.62%	0.45%	0.62%	0.51%	0.25%	0.38%	0.22%	0.15%	0.49%
P1	2.17%	2.19%	1.92%	1.89%	75.30%	1.80%	1.51%	1.40%	1.21%	1.14%	1.33%	1.60%	0.95%	0.89%	0.73%	0.56%	0.71%	0.70%	0.21%	0.48%	0.27%	0.15%	0.59%
PHA stimulated PBMC																							
V1	1.43%	1.56%	1.35%	1.54%	84.26%	1.13%	0.95%	0.92%	0.68%	0.82%	0.73%	0.85%	0.72%	0.57%	0.43%	0.32%	0.31%	0.45%	0.11%	0.22%	0.12%	0.08%	0.44%
V2	1.56%	1.65%	1.47%	1.49%	82.35%	1.37%	0.96%	1.07%	0.82%	0.94%	0.81%	0.95%	0.88%	0.67%	0.49%	0.40%	0.43%	0.49%	0.18%	0.30%	0.15%	0.15%	0.44%
V3	1.54%	1.82%	1.56%	1.58%	82.35%	1.29%	1.11%	1.03%	0.88%	0.82%	0.82%	0.93%	0.84%	0.59%	0.49%	0.34%	0.29%	0.53%	0.13%	0.26%	0.16%	0.11%	0.50%
V4	1.40%	1.70%	1.45%	1.56%	83.34%	1.20%	1.01%	1.05%	0.81%	0.87%	0.81%	0.83%	0.70%	0.59%	0.43%	0.37%	0.29%	0.47%	0.17%	0.22%	0.15%	0.08%	0.48%
S	1.84%	1.76%	1.48%	1.55%	81.98%	1.27%	1.11%	1.04%	0.77%	0.87%	0.84%	0.98%	0.86%	0.66%	0.51%	0.33%	0.37%	0.49%	0.16%	0.34%	0.19%	0.13%	0.45%
P5	1.61%	1.80%	1.29%	1.39%	83.30%	1.19%	1.01%	0.93%	0.78%	0.91%	0.83%	1.05%	0.78%	0.55%	0.54%	0.24%	0.34%	0.46%	0.13%	0.29%	0.12%	0.08%	0.40%
P4	2.04%	2.12%	1.61%	1.40%	80.33%	1.68%	1.23%	1.17%	0.87%	1.04%	0.94%	0.95%	0.72%	0.66%	0.69%	0.35%	0.37%	0.56%	0.17%	0.32%	0.18%	0.10%	0.52%
P3	1.67%	1.87%	1.42%	1.47%	82.54%	1.39%	1.13%	1.08%	0.71%	0.88%	0.89%	0.93%	0.78%	0.54%	0.57%	0.32%	0.29%	0.43%	0.12%	0.25%	0.15%	0.09%	0.47%
P2	2.28%	2.02%	1.71%	1.48%	79.62%	1.54%	1.13%	1.12%	0.89%	1.07%	0.95%	1.10%	0.83%	0.72%	0.72%	0.41%	0.55%	0.42%	0.23%	0.33%	0.17%	0.19%	0.51%
P1	2.35%	2.19%	1.92%	1.66%	76.98%	1.81%	1.42%	1.26%	1.03%	1.10%	0.98%	1.27%	0.94%	0.91%	0.75%	0.48%	0.66%	0.55%	0.26%	0.45%	0.21%	0.20%	0.62%
Lymphoblastoids																							
V1	1.92%	2.19%	1.91%	1.94%	77.80%	1.65%	1.34%	1.41%	1.03%	1.02%	1.20%	1.12%	1.11%	0.79%	0.61%	0.40%	0.41%	0.70%	0.20%	0.32%	0.20%	0.12%	0.60%
V2	2.37%	2.42%	1.96%	1.88%	76.39%	1.88%	1.59%	1.37%	0.97%	1.20%	1.33%	1.16%	0.86%	0.92%	0.64%	0.49%	0.57%	0.55%	0.20%	0.39%	0.16%	0.14%	0.56%
V3	2.09%	2.70%	2.12%	2.21%	75.49%	1.96%	1.63%	1.49%	1.01%	1.18%	1.23%	1.31%	1.02%	0.75%	0.60%	0.48%	0.40%	0.75%	0.11%	0.37%	0.24%	0.09%	0.76%
V4	1.82%	2.21%	1.73%	1.91%	78.88%	1.62%	1.48%	1.38%	0.86%	1.10%	1.15%	1.01%	0.92%	0.71%	0.53%	0.35%	0.35%	0.62%	0.16%	0.27%	0.20%	0.08%	0.64%
S	2.31%	2.33%	1.97%	1.81%	76.46%	1.58%	1.41%	1.39%	1.03%	1.29%	1.32%	1.25%	1.04%	0.84%	0.65%	0.58%	0.58%	0.64%	0.18%	0.40%	0.23%	0.18%	0.56%
P5	2.03%	2.09%	1.80%	1.77%	78.70%	1.51%	1.25%	1.34%	1.04%	1.24%	1.40%	0.94%	0.79%	0.83%	0.68%	0.41%	0.35%	0.58%	0.16%	0.33%	0.18%	0.09%	0.52%
P4	2.37%	2.82%	2.16%	2.04%	74.39%	2.09%	1.65%	1.57%	1.04%	1.35%	1.32%	1.35%	0.92%	0.82%	0.78%	0.37%	0.40%	0.52%	0.26%	0.36%	0.22%	0.09%	0.57%
P3	1.88%	2.24%	1.80%	1.84%	79.40%	1.60%	1.31%	1.25%	0.89%	0.96%	1.10%	1.08%	0.88%	0.68%	0.57%	0.37%	0.40%	0.58%	0.10%	0.27%	0.15%	0.07%	0.66%
P2	2.12%	2.25%	1.89%	1.62%	76.98%	1.90%	1.43%	1.33%	1.15%	1.18%	1.18%	1.27%	0.90%	0.85%	0.75%	0.42%	0.52%	0.69%	0.20%	0.45%	0.19%	0.17%	0.62%
P1	3.10%	3.22%	2.50%	2.20%	69.33%	2.22%	1.79%	1.71%	1.36%	1.51%	1.68%	1.72%	1.22%	1.06%	0.91%	0.74%	0.83%	0.85%	0.30%	0.55%	0.29%	0.21%	0.72%

Table 2: (suppl.) Proportion of reads per chromosome. The reads were counted on every chromosome. In this table, the percentage of reads mapped on every chromosome is reported for every cell type and every viewpoint. At least 70% of the reads are localized on the chromosome 5. The non-digested product was excluded from this analysis.

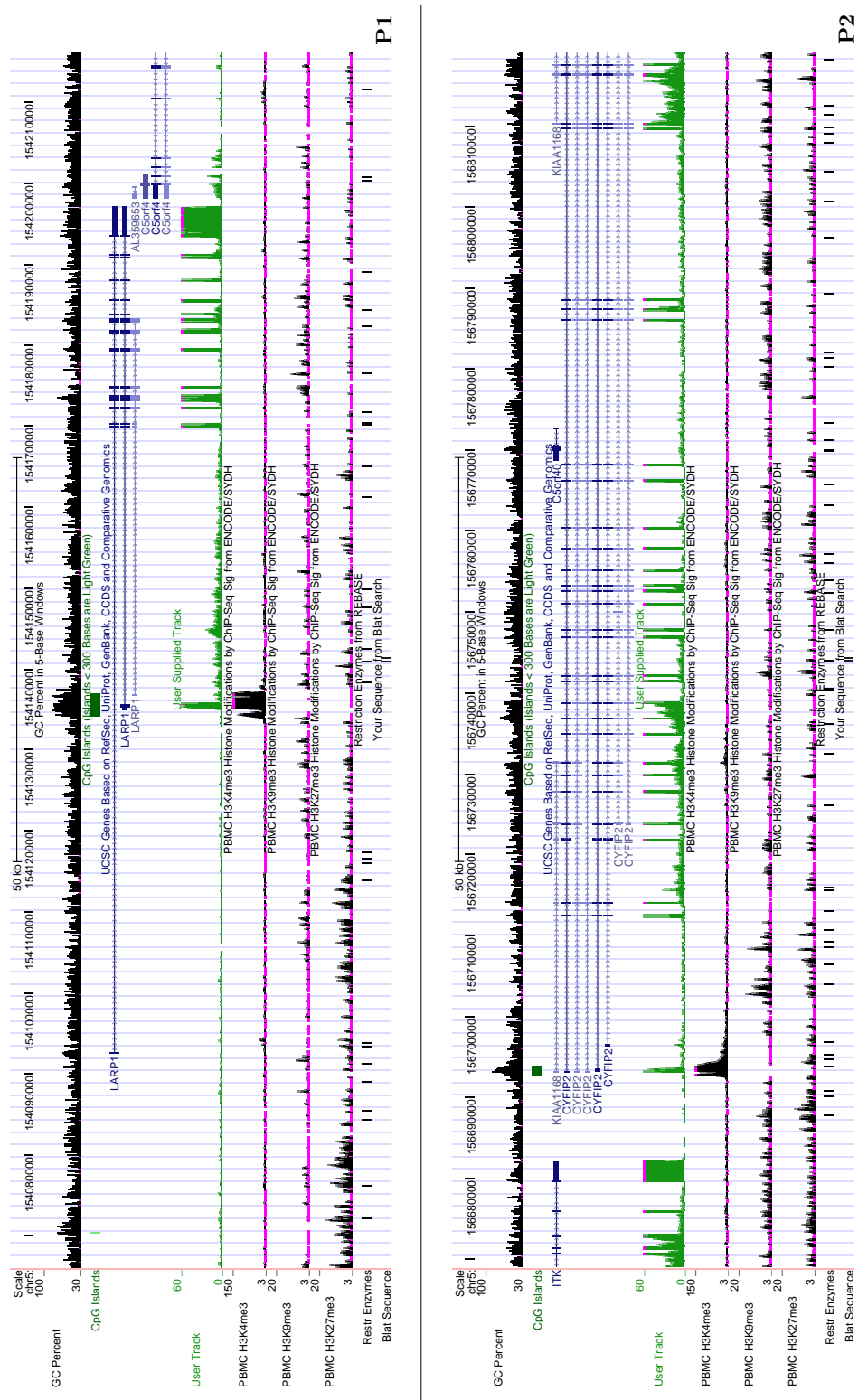
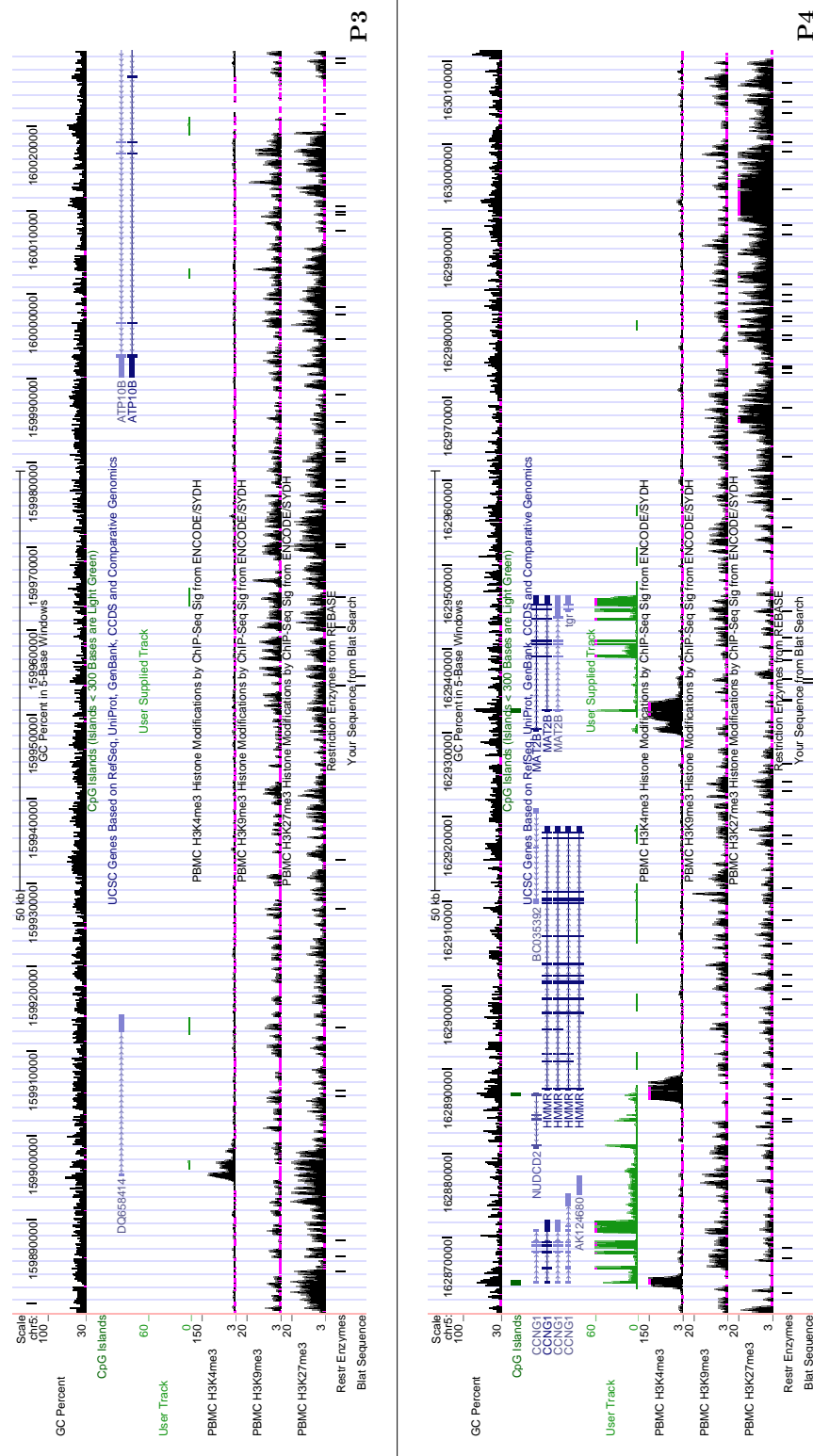
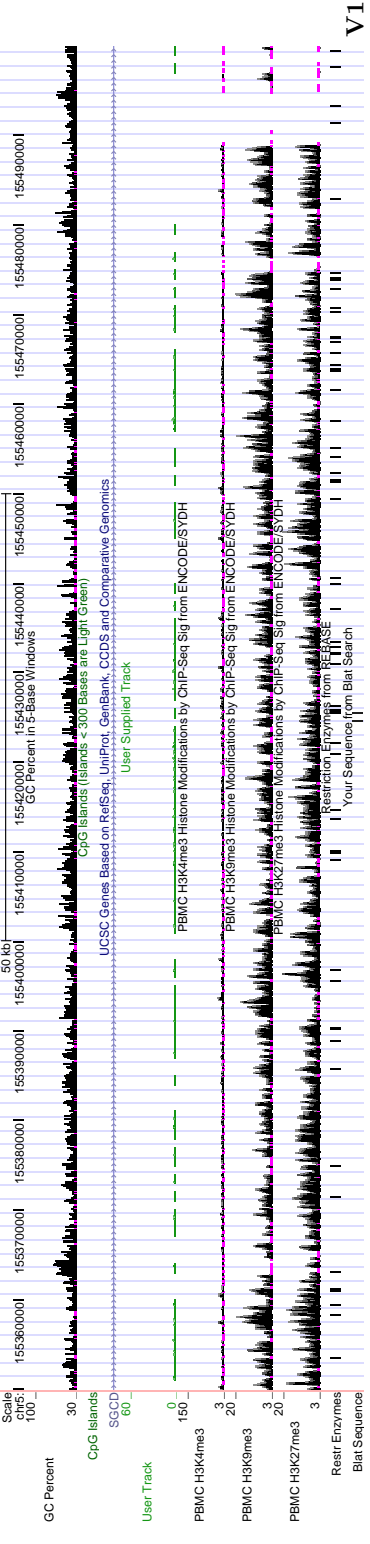
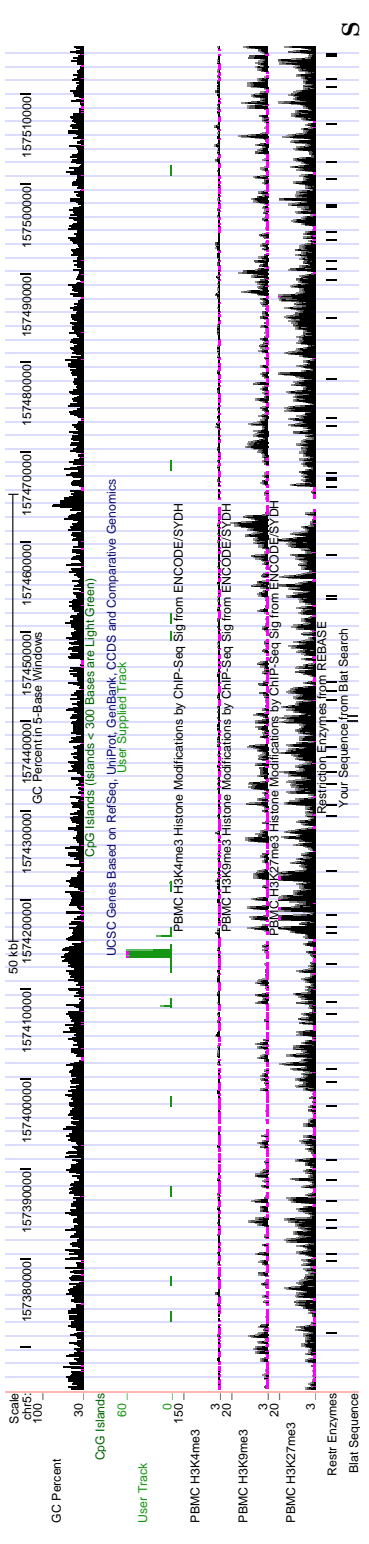
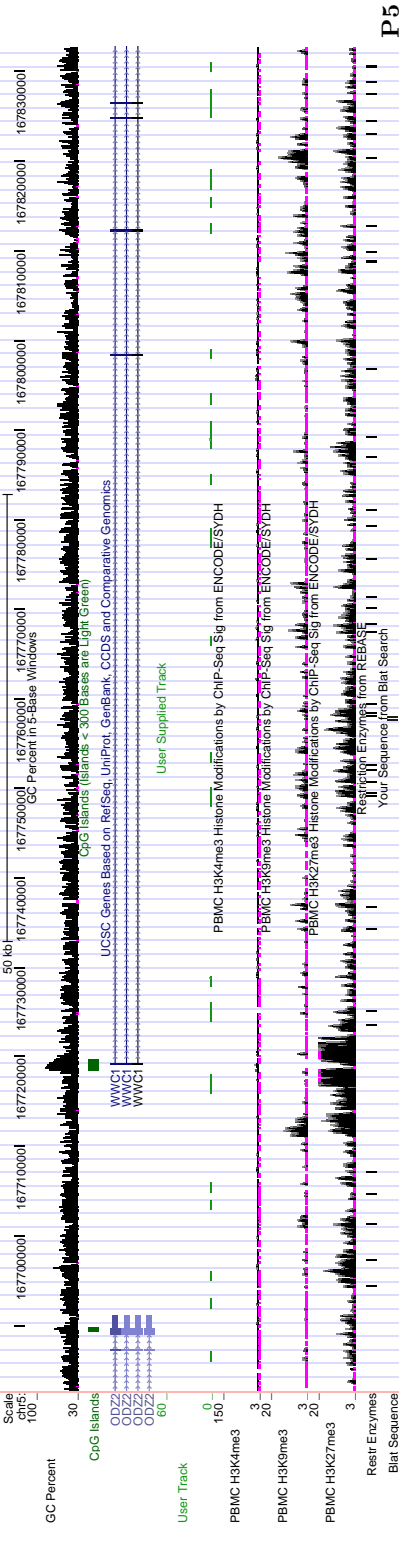


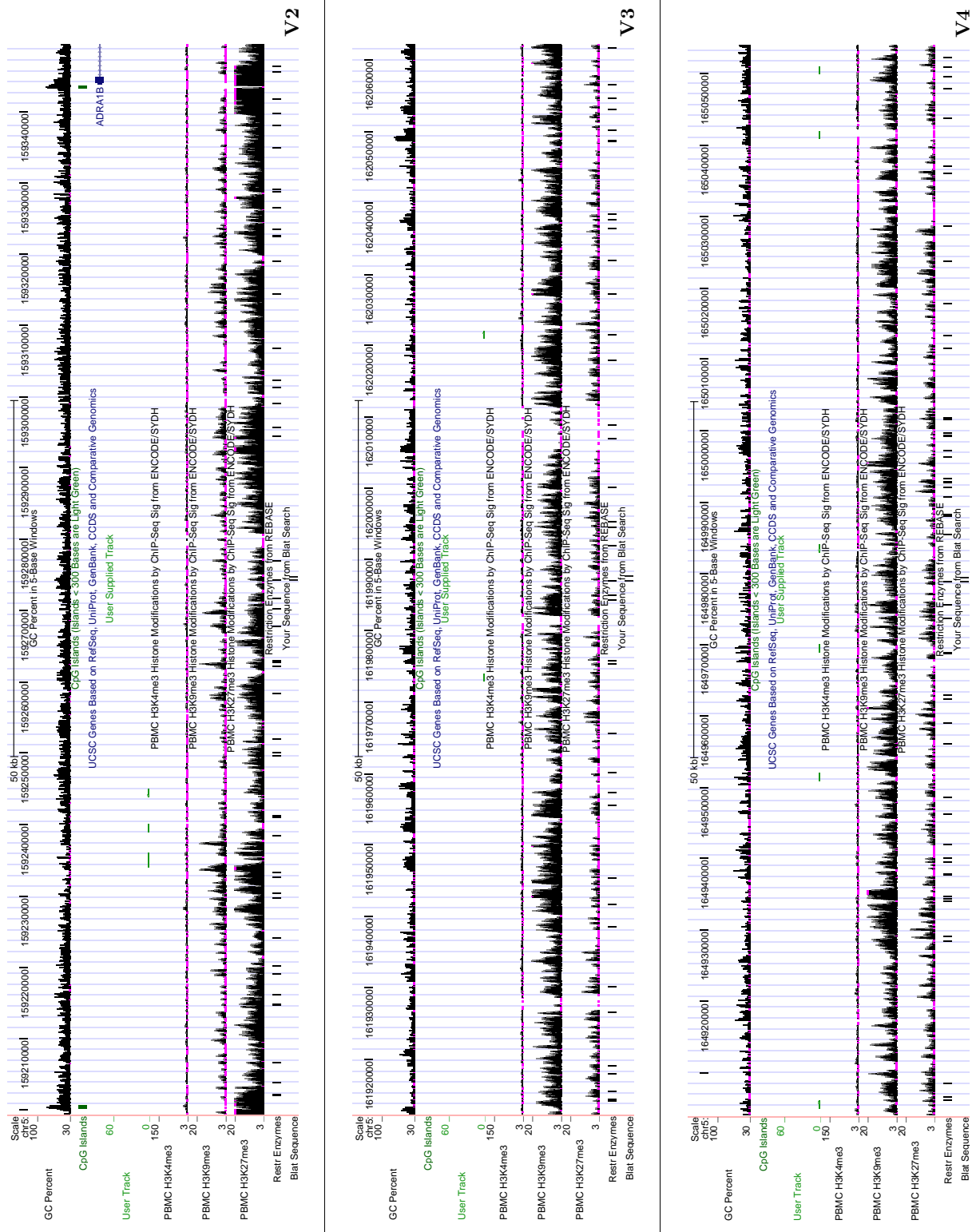
Figure 36: (suppl.) Local chromatin properties around the ten selected viewpoints in PBMC. Using the UCSC genome-browser, we analyzed in PBMC the local chromatin properties in 150kb intervals around selected viewpoints. From the top to the bottom: the GC content (5bp window, hg19_gc5Base); the presence of CpG island (hg19_cpgIslandExt); the UCSC genes (hg19_knownGene), mRNA expression in PBMC cells (display as user supplied track, GEO GSM669616); histones modifications in PBMC



continued — (H3K4me3 wgEncodeSydhHistonePbmcH3k04me3bUcdSig, H3K9me3 wgEncodeSydhHistonePbmcH3k09me3bUcdSig, and H3K27me3 wgEncodeSydhHistonePbmcH3k27me3bUcdSig); the HindIII restriction sites. The last track corresponds to the location of the two 4C-Seq primers. While 4C mapping was performed on the hg18 human assembly, these UCSC genome browser outputs correspond to the hg19 assembly in which chromatin data are available for PBMC cells. Lift-over tool was used to adapt 4C primers coordinates.



continued



continued

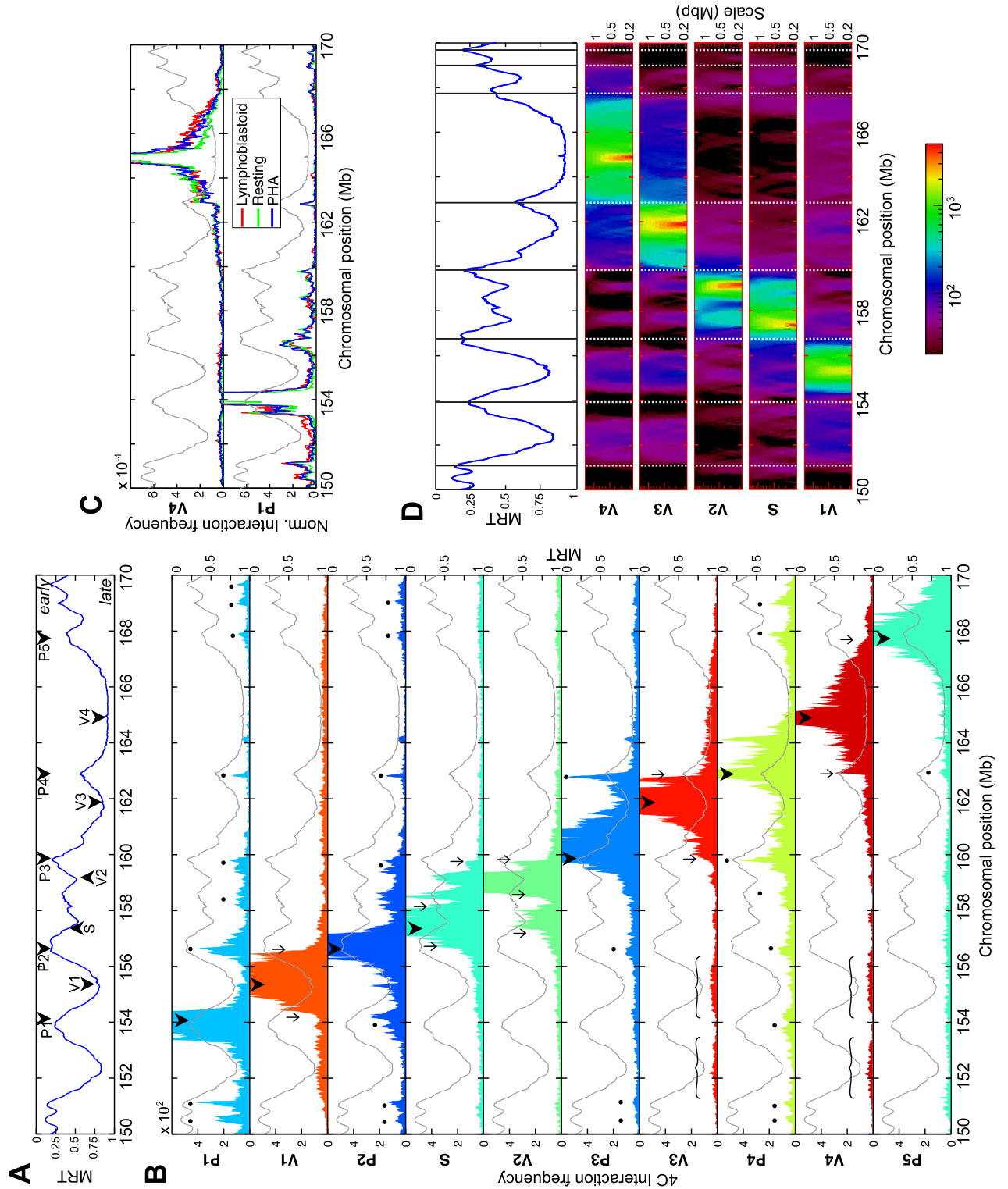
are located on chromosome 5 and a much lower read density was observed in other chromosomes (table 2), which is in agreement with the existence of chromosome territories. Similar proportions of intra-chromosomal interactions were observed for the other cell types used in the study. Within chromosome 5, most contacts were local and the interaction frequencies rapidly dropped with increasing genomic separation (table 1), as previously reported [de Wit and de Laat, 2012].

4.2.1- Early timing peaks delineate structural self-interacting domains and preferentially contact each other

4C data were locally averaged using a running mean within a window of 20 HindIII sites (~100kb). The plots show that interaction frequencies of the valley viewpoints (V1, V2, V3 and V4), as well as of the S viewpoint, quickly decrease to a few percents of the maximum signal over a few Mb distance from the considered viewpoint (figure 37B). This decrease does not follow a simple pattern; indeed, sharp drops can be seen. Strikingly, most of these drops seem to be localized at, or close to, timing peaks (figure 37B), arrows). This suggests that contacts between late-replicating regions are somehow limited by intervening timing peaks. To further illustrate that these sharp drops in the 4C profiles really, we displayed 4C data using a 2D space-scale representation. Since 4C raw signals present strong deviations from the local mean, we represented the median value in sliding windows of size ranging from 200 kb to 1.35 Mb. The resulting images were obtained by encoding the median values with a logarithmic color map (figure 37D). It became apparent that 4C signal amplitudes are larger within a domain bordered by two timing peaks, and that clear transitions occur at timing peaks. This analysis confirmed that timing peaks delineate frontiers limiting the DNA contacts of neighboring late replicating regions, and thus sheds light on the existence of self-interacting chromosomal domains bordered by timing peaks. Interestingly, beyond this local regime, late-replicated loci preferentially interact with other late replicating regions, as can be seen for the long-range interactions of the V3 and V4 valley viewpoints (figure 37B, curly brackets).

The interaction frequencies obtained with viewpoints localized on timing peaks (P1 to P5) display profiles strongly different from those of the valley-viewpoints. P1, P2 and P4 (and to a lesser extent P3) profiles exhibit preferential contacts with nearly every neighboring timing peak, even with those located 10 Mb away (figure 37B, dots), showing that they loop out the intervening late replicating domains to contact each other. Although localized on an early peak, the P5 viewpoint does not display specific contacts with other timing peaks,

Figure 37: 4C interaction profiles associated with replication timing in lymphoblastoid cells. A- Mean replication timing (MRT) of the lymphoblastoid cell line Gm06990, in the portion 150-170 Mb of human chromosome 5 (Hansen et al. 2010). Early (resp. late) replicating sequences correspond to MRT values 0 (resp. 1). The arrowheads indicate the position of the viewpoints used for the 4C analysis. B- Interaction frequencies profiles of the ten viewpoints. For each panel, the light grey curve corresponds to the replication timing profile (right y-axis) and the colored filled curve to the 4C interaction frequency profile (left y-axis). The arrowheads indicate the position of the viewpoints; the thin arrows indicate sharp drops in interaction frequency coinciding with early timing peaks; the curly brackets indicate partners of V3 and V4 viewpoints; and the dots indicate preferential partners of P viewpoints coinciding with timing peaks.



continued — C- 4C interaction frequency is highly reproducible from one cell line to another. To normalize for differences in read numbers, the running mean value (B) has been divided by the total number of reads mapped on chromosome 5 (red = lymphoblastoid, green = resting PBMC, blue = PHA-stimulated PBMC). Top, V4 viewpoint; bottom, P1 viewpoint. D- Timing peaks correspond to zones of sharp changes in 4C data interaction frequency. (Top) The position of timing peaks are marked along the MRT profile in Gm06990 (vertical bars). (Middle) 2D space-scale representation of the 4C interaction frequencies of S- and V-viewpoints. Each horizontal line corresponds to the median value of the interaction frequency computed in sliding windows of size ranging from 200 kb to 1.35 Mb. Due to the very large median-value dynamics, a logarithmic colormap was used (bottom).

which might be due to specific features of this viewpoint (see Discussion). Interestingly, whereas the contacts between timing-valleys are spread over Mb (figure 37B, curly brackets), the contacts between early timing peaks are restricted to shorter distances (figure 37B, dots), suggesting the existence of a stronger and more specific anchoring mechanism between early timing peaks than between late regions.

To address whether this replication-related 3D-organization of chromatin is also found in non-cycling cells, we performed similar analyses in resting PBMC, and, as a control, in cycling phytohaemagglutinin-stimulated PBMC (figure 38. For both cell types, we used the replication timing of Gm06990 cell line as a reference since the timing profile in the region containing the viewpoints is remarkably similar in different cell types (figure 25A in chapter 3). The local 4C profiles obtained with the 3 different cell types remarkably superimpose on each other (figure 37C), arguing for the existence of a robust and common 3D chromatin conformation. In resting and cycling PBMC, we also observed frontiers localized at timing peaks that prevent the contacts of neighboring late-replicated regions (figure 37C and figure 39). In addition, using timing peaks as viewpoints, we noticed preferred contacts with every timing peak. This not only confirms the preferential contacts between early-replicated regions as previously observed [Yaffe et al., 2010], but more strikingly also shows that the contact frequency profile is strongly correlated with the timing profile in the three related cell types. Moreover, the fact that the local conformation in resting PBMC is similar to the one in cycling cells implies that this correlation exists in G0 and does not result from ongoing DNA replication. It suggests that this organization pattern is conserved all along the cell cycle.

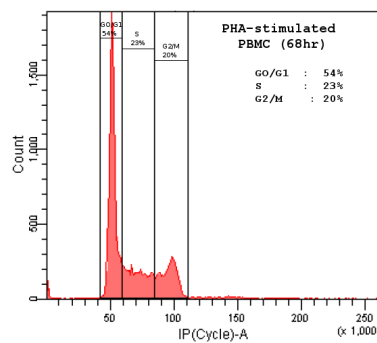


Figure 38: (*suppl.*) **Entry of PHA-stimulated cells into cell cycle.** An aliquot of PHA-stimulated PBMC used in 4C experiments was prepared for PI/FACS analysis. DNA content of cells was analyzed using FACSDiva™. The percentage of cells in G0/G1, S and G2/M is shown. Compared to resting cells, 40% of the cells are in S or G2/M phase after a 68hr-PHA treatment. This attests for an efficient cell-cycle entry.

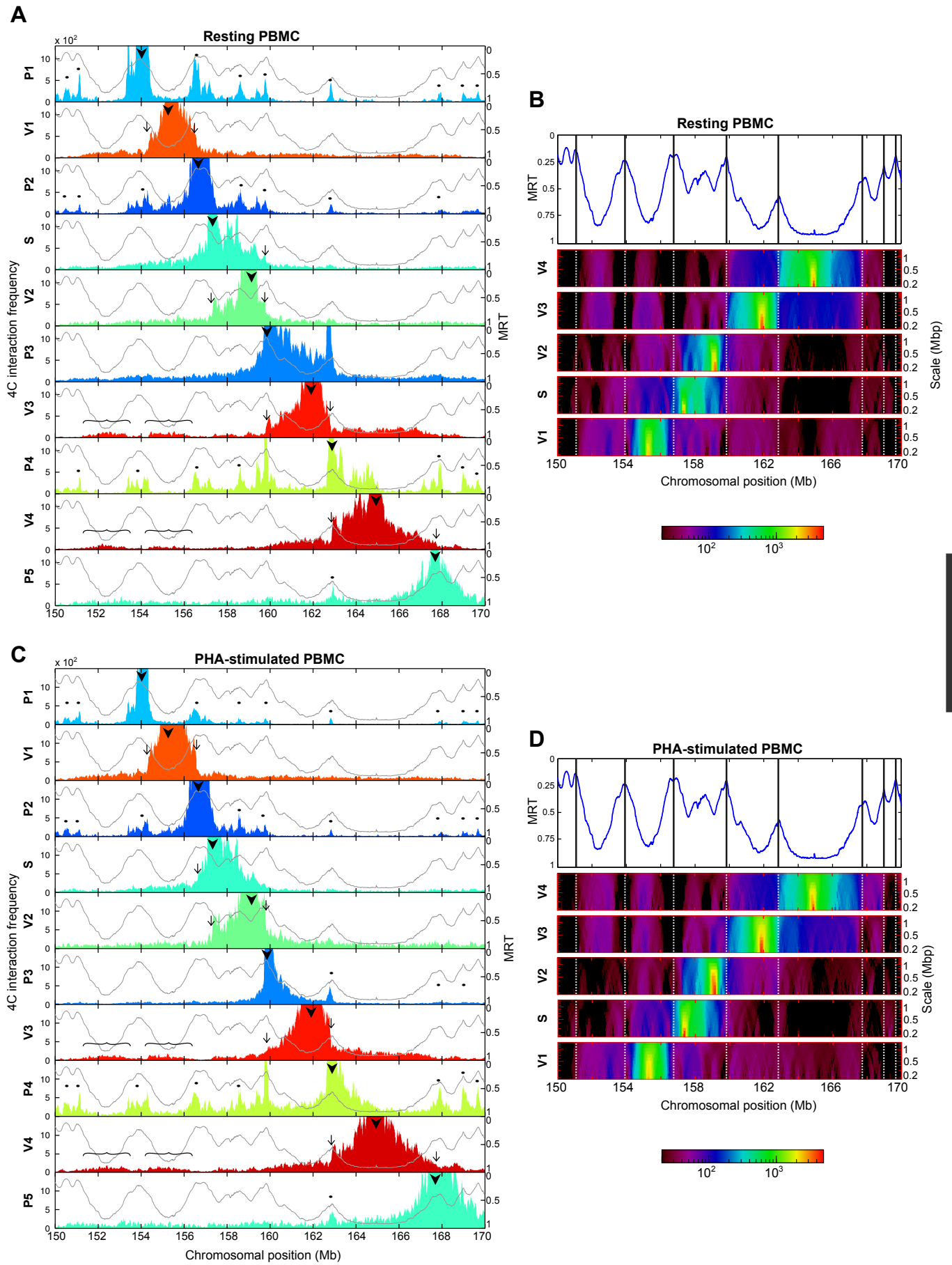


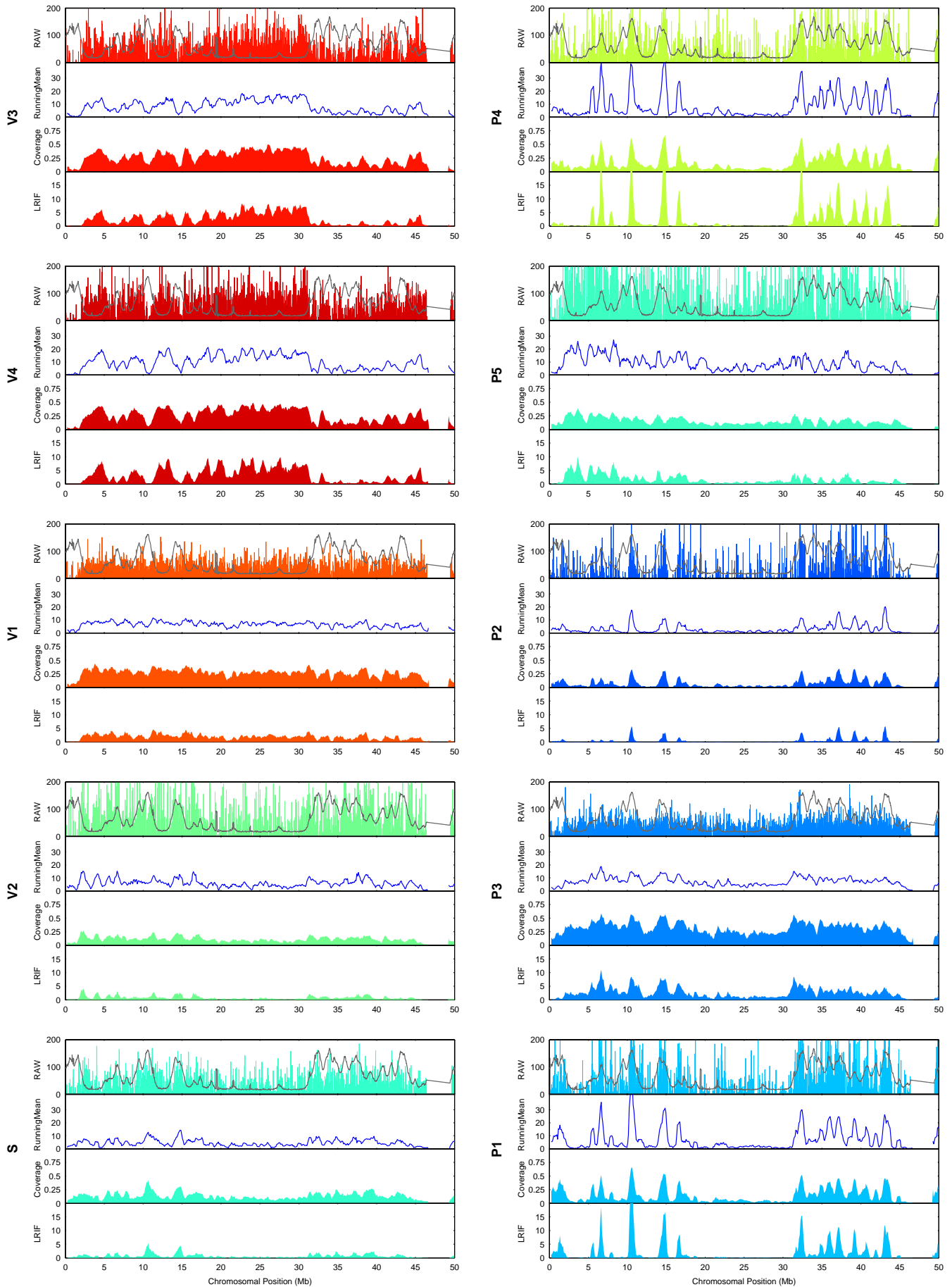
Figure 39: (suppl.) Same as in figure 37, but in PBMC (A, B) and PHA-stimulated PBMC (C, D)

4.2.2- Long-range interaction frequency profiles faithfully reproduce the replication timing profile

In 4C experiments, the measured interaction frequencies have been shown to rapidly drop with increasing genomic distance [Simonis et al., 2007]. As a consequence, only few informative ligation events can be captured while analyzing long-range contacts. To collect the long-range contacts, it has been shown that the contribution of isolated reads improperly amplified by PCR should be diminished, and the contribution of low-abundance densely packed reads reinforced [Simonis et al., 2007]. Accordingly, we defined a variable called Long-Range Interaction Frequency (LRIF, see Methods), that operates such a filtering. While the LRIF signal is synchronous with the coverage and the running mean defined at 500 kb (figure 40), it offers better signal to noise ratio, facilitating subsequent analyses.

All along the chromosome, the long-range contacts obtained from V1, V3, V4 on the one hand, and from P1, P2, P3, P4 on the other hand, display opposed oscillations in the three cell types tested, i.e. P viewpoints maxima coincided with V viewpoints minima and *vice versa* (figure 41A,B 42). To quantify this observation for each viewpoint on the whole chromosome 5, we detected long-range contacts by local maxima in the LRIF signal, and measured the pairwise distances separating each local maximum from the closest local maximum of another viewpoint. As an illustration of the results, the histogram of distances between P4 and P1 peaks is centered at zero, indicating that the long-range contacts of P1 and P4 viewpoints overlap and that P1 and P4 viewpoints indeed interact with the same partners (figure 41D). On the contrary, the V4-P1 distance distribution is clearly depleted around zero, showing that V4 and P1 have different distant partners (figure 41D). To depict all the pairwise combinations, we generated a heat-map in which a green-red colormap quantifies overlap of long-range contacts identified from LRIF signals (figure 41E). The unbiased hierarchical clustering performed revealed the existence of three groups. The first one corresponds to viewpoints on timing peaks P1, P2, P3 and P4 that display LRIF signal in opposition of phase with the second group that includes valley viewpoints V1, V3 and V4. The third group corresponds to V2, S and P5 viewpoints that display intermediate 4C pro-

Figure 40 : (suppl.) The detection of long-range interactions is independent on the variable choice. Since long-range interactions are rare events in a 4C library, we defined three different variables counting interaction events in a running window of size of 500 kb. The three different variables are shown for resting PBMC in the first 60Mb of human chromosome 5. These three variables display synchronized maxima and thus highlight similar long-range interactions. From top to bottom: (1) the RAW signal. Each bar corresponds to the number of reads per restriction fragment. The light grey curve corresponds to the replication timing (2) The Running Mean of the RAW signal on all the HindIII restriction size of the genome. This running mean was calculated in a sliding window of 500 kb. At each step, the window slides of 100kb. (3) The coverage. We measured the percentage of HindIII restriction sites covered by reads. If a strong interaction exists, most of the restriction sites of the interacting partner should be covered by reads. Close to the viewpoints, the coverage is 1; farther away the coverage oscillates between 0 and 1. This value should reinforce the importance of the rare ligation events (few reads) concentrated on a small portion of chromatin. (4) To distinguish the case of concentrated ligation events with few reads from those with large read numbers, we defined a Long-Range Interaction Frequency (LRIF) variable corresponding to the product Coverage x Mean in a sliding window of 500 kb. This variable exhibits a large contrast signal:noise facilitating the detection of maxima (and thus of long-range interactions).



files; the hierarchical clustering performed yet tends to position them closer to early viewpoints than to late ones. Similar results were obtained with the three cell lines used in this study, showing the remarkable conservation of the long-range interactions frequencies (figure 41C,E). Taken together, we conclude that chromosomal domains with opposite replication timing tend to segregate in nuclear space, and they do so even in non-dividing quiescent cells.

We then asked to which extent the long-range partners of every viewpoint coincide with timing peaks. We therefore measured, for each viewpoint and on the whole genome, the distances between every 4C peaks and the closest timing peak. The obtained histograms for lymphoblastoid cells (figure 43A) clearly show that the long-range contacts of the P4 viewpoint, for example, are close to timing peaks, as the histogram is centered on zero. On the contrary, the histogram for the V3 viewpoint is depleted in values around zero indicating that V3 partners are away from timing peaks. V2 viewpoint exhibits an intermediate situation: the distribution histogram is neither peaked nor depleted around zero. To depict this information for all viewpoints and the three cell lines, a green-red colormap was used to show the coincidence between timing peaks and long-range contacts (figure 43B). Our results show that the long-range partners of P1, P2, P3 and P4 viewpoints overlap with timing peaks for the whole genome. Conversely, V1, V3 and V4 do not interact with timing peaks. Overall, we observed that the long-range interaction frequency profiles not only correlate genome-wide with replication timing, but also strikingly reproduce the replication timing signal over very large distances along the chromosome 5. Interestingly, similar tendency can still be observed on other chromosomes (figure 44), although with a much lower interaction strength (about two order of magnitude). The fact that the interaction frequency profiles tightly follow the replication timing strongly argues in favor of a replication imprint on chromatin organization.

Figure 41 : Chromosomal domains with opposite replication timing segregate in the nuclear space.

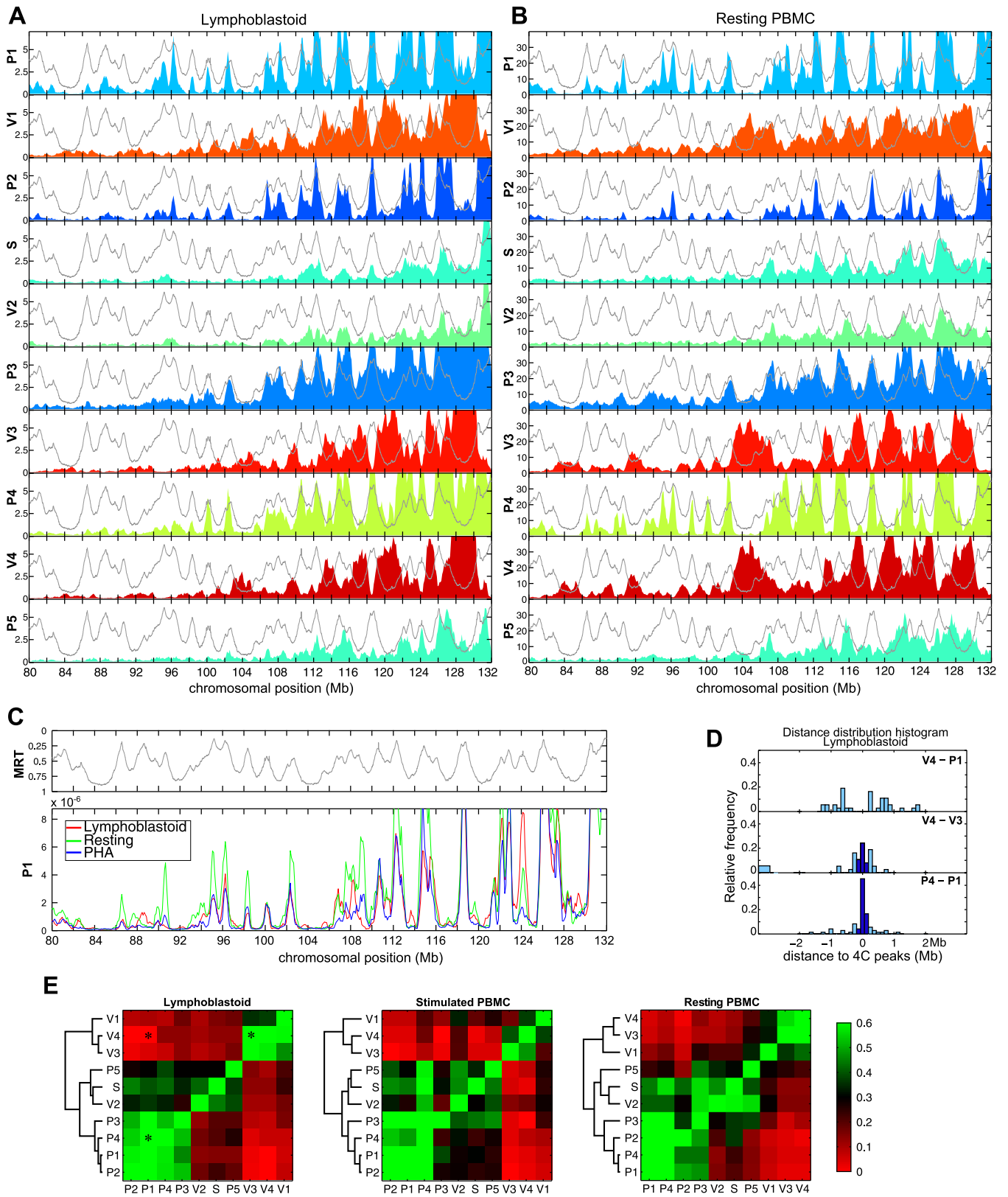
A- Long-range interaction frequency (LRIF) in lymphoblastoid cells for the ten viewpoints (see Text and Methods for the definitions). The colored filled curves correspond to the 4C profiles in a region of chromosome 5 centered 60 Mb upstream of the viewpoints. The light grey lines correspond to the replication timing in lymphoblastoid cells (Hansen et al., 2010) plotted with the same y-axis as in figure 37A.

B- Similar analysis performed with resting PBMC.

C- Comparison of LRIF values of the P1 viewpoint in lymphoblastoid cells (red), resting PBMC (green) and PHA-stimulated PBMC (blue). LRIF values have been normalized for differences in read numbers. The upper panel corresponds to the replication timing in lymphoblastoid cells.

D- Distribution of distances between the 4C peaks detected in the LRIF profiles. We detected the 4C peaks and measured the distances between each of them and the closest 4C peak obtained with another viewpoint. All pairwise distances were measured over the whole chromosome 5. Three distance distribution histograms are shown here as examples (V4-P1, V4-V3 and P4-P1). A distribution centered on zero (as P4-P1) indicates that the LRIF maxima of the two viewpoints overlap.

E- To quantify the similarities between LRIF signals from different viewpoints, we calculated the percentage of occurrences where the pairwise distances measured in D are smaller than 250 kb. This percentage was encoded with a green-red color map. Similar signals are green (V4/V3 and P4/P1), whereas out of phase signals are red (V4/P1). The asterisks correspond to the distribution histograms shown in D.



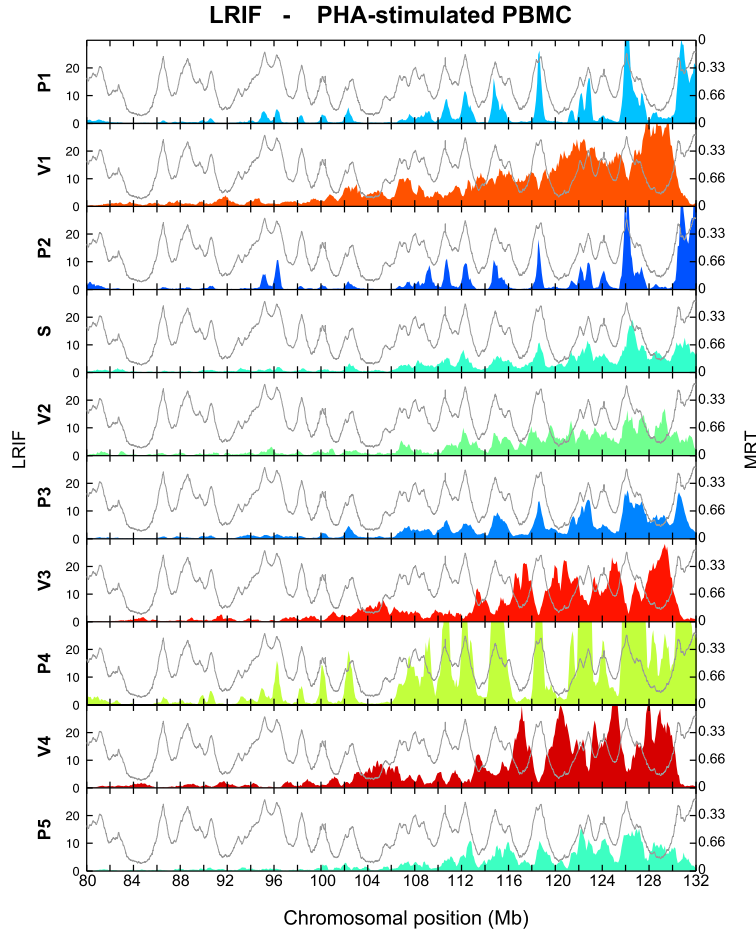


Figure 42: (suppl.) LRIF signals in PHA-stimulated PBMC. LRIF values in 68hr-stimulated PBMC for the ten viewpoints (see Text and Methods for the definitions). The colored filled curves correspond to the 4C profiles and the light grey curves correspond to the replication timing.

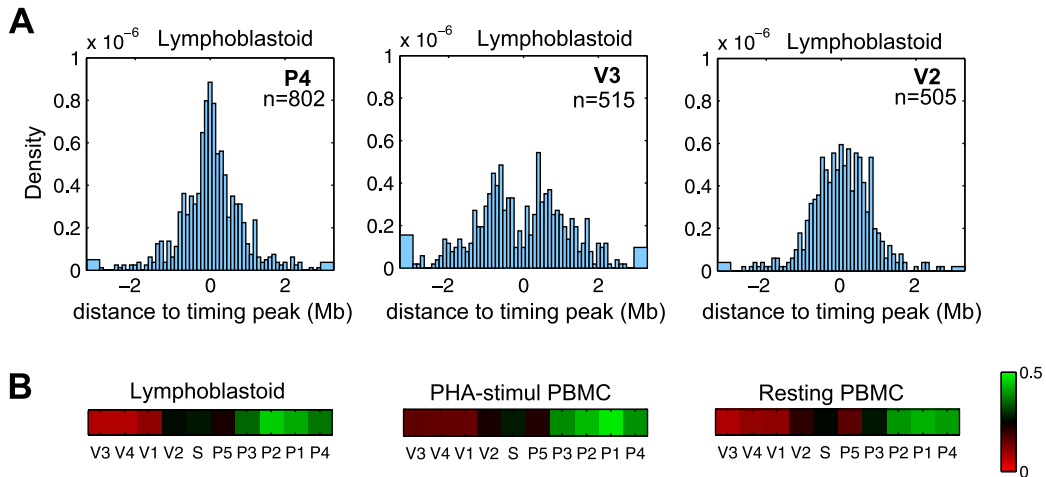


Figure 43: Viewpoints localized on local timing peaks mostly interact with other timing-peaked regions. A- Histograms of the distance between every 4C peak and the closest timing peak on the whole genome. 4C peaks were detected in the whole genome, then we measured the distance between each of them and the closest timing peak. The three histograms represent these distances measured in lymphoblastoid cells for P4 (left), V3 (middle) and V2 (right) viewpoints. A distribution centered on zero (P4) indicated that the 4C peaks are close to timing peaks. n = number of 4C peaks detected. B- We computed the percentage of distances smaller than 250 kb and represented this value with a green-red color map. 4C signals in phase with the replication timing are green (P4) whereas anti-correlated signals are red (V3).

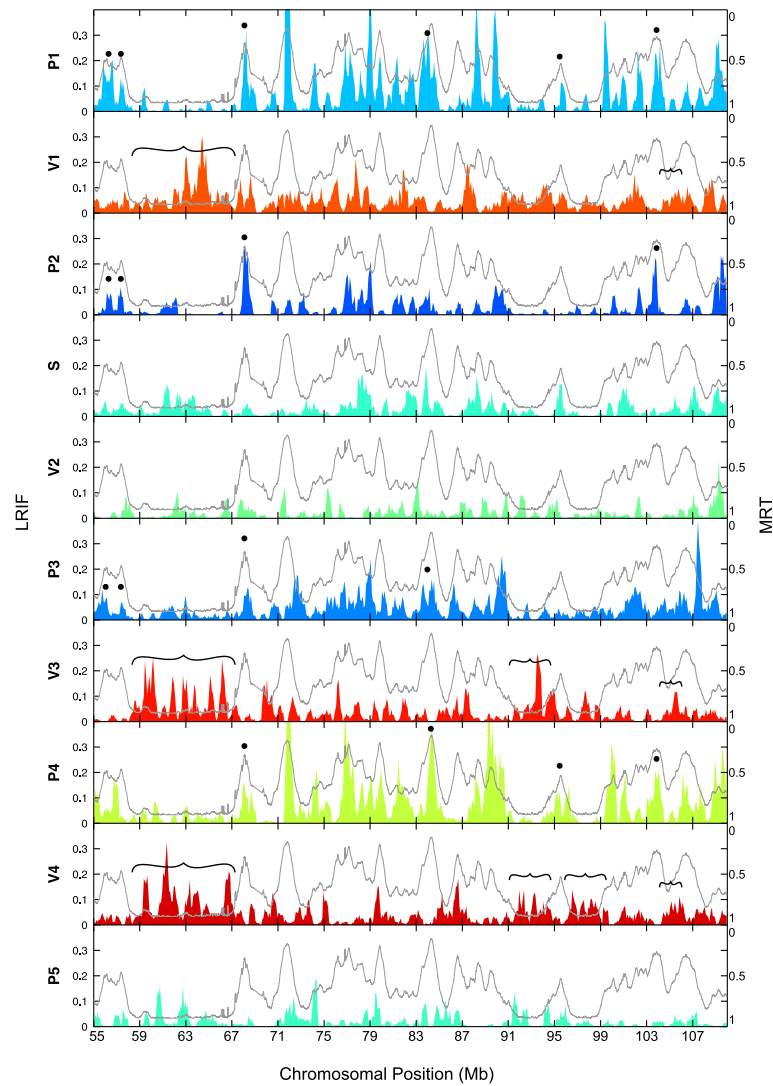


Figure 44: (suppl.) Inter-chromosomal LRIF signals in resting PBMC. Long-range interaction frequency (LRIF) in resting PBMC for the ten viewpoints over 55 Mb of the human chromosome 4. The colored filled curves correspond to the 4C profiles and the light grey curves correspond to the replication timing. Note that, as compared to figure 41, the LRIF y-axis has been considerably rescaled by a factor 100 to allow the visualization of small LRIF fluctuations. Even when considering different chromosomes, the late-replicating V1/V3/V4 viewpoints mostly interact with other late-replicated loci (brackets), whereas the early ones (P1, P2, P3, P4) contact other early timing peaks (dots).

4.2.3- *In-situ* distance measurements confirm the physical proximity between loci corresponding to early timing peaks.

We used fluorescent *in-situ* hybridizations in PBMC to confirm that early-regions preferentially interact with each other. We selected BAC probes localized in the 150-170 Mb portion of human chromosome 5. Using three colors FISH, we measured three different sets of physical distances, a first one (EE) corresponding to distances between two local timing-peaks, a second set (LL) corresponding to distances between two timing-valleys, and a third set (LE) corresponding to distances between a peak and a valley (figure 45A). Results can be depicted by cumulative histograms. For example, as predicted by the 4C results, the d5 (EE) physical distance is smaller than d17 (LE) even if both correspond to a 7.6 Mb genomic length (figure 45C(c)). Similarly, the d3 physical distance (EE) is slightly smaller than d1 (LE) and d2 (LE) although d3 corresponds to a much larger DNA fragment than both d1 and d2 (figure 45C(a)). Likewise, d28, d29, and d30 present similar physical distances although d30 (EE) corresponds to a genomic length twice larger than the LE distances d29 and d30 (figure 45C(b)). From these three examples, we concluded that d3, d5 and d30, which correspond to physical distances between timing-peaked regions, are smaller than expected if the genomic separation was the only parameter determining the 3D-distance in the nucleus.

It is established that the distance histograms can be easily fitted by a Rayleigh law (figure 45D) that only depends on a single parameter σ (Methods). This Rayleigh parameter perfectly describes the shape of the obtained histograms and can be used to define, for each probe combination, the corresponding Mean Square Physical Distance [Yokota et al., 1995]. We plotted the Mean Square Physical Distance as a function of the genomic separation between the probes (figure 45E). Among the dots deviating from the mean tendency, most correspond to peak-to-peak distances (d3, d30, d5, d18, and to a lesser extent d6). In particular, d3 is the smallest physical distance, in agreement with 4C results showing that the strongest interaction in PBMC occurs between P3 and P4 peaks (figure 39). This indicates that, for the same genomic separation, the physical distance between two early replicated loci tend to be smaller than the distance between a late and an early replicated locus. Nevertheless, some early-to-early physical distances are not smaller than the mean tendency (d4/d11, d12, and d15). Therefore, if 4C-contacts are not necessarily seen by FISH, the FISH results largely agree with those obtained in 4C experiments.

We did not detect spatial proximity between probes in timing-valleys (figure 45E). This may be explained by the fact that contacts between timing-valleys are spread over the width of the valleys (figure 39A, curly brackets), suggesting a more relax interaction pattern between late-replicated regions than between timing peaks.

The 4C profiles in PBMC indicate that timing peaks are found at frontiers limiting the contacts of neighboring late-replicated regions (figure 39A). To test whether this is also observed by FISH, we performed measurements around the P4 timing peak. For comparable genomic separation, physical distances measured across the timing peak tend to be larger than those performed in the valley (figure 45F, middle). This tendency is supported by the frequency of colocalization between the probes used above in the distance measurements (figure 45F, bottom). Probes separated by the timing peak less frequently colocalized than probes located in the same timing valley. These observations strengthen the conclusion that a timing peak can limit the interaction capabilities of a late-

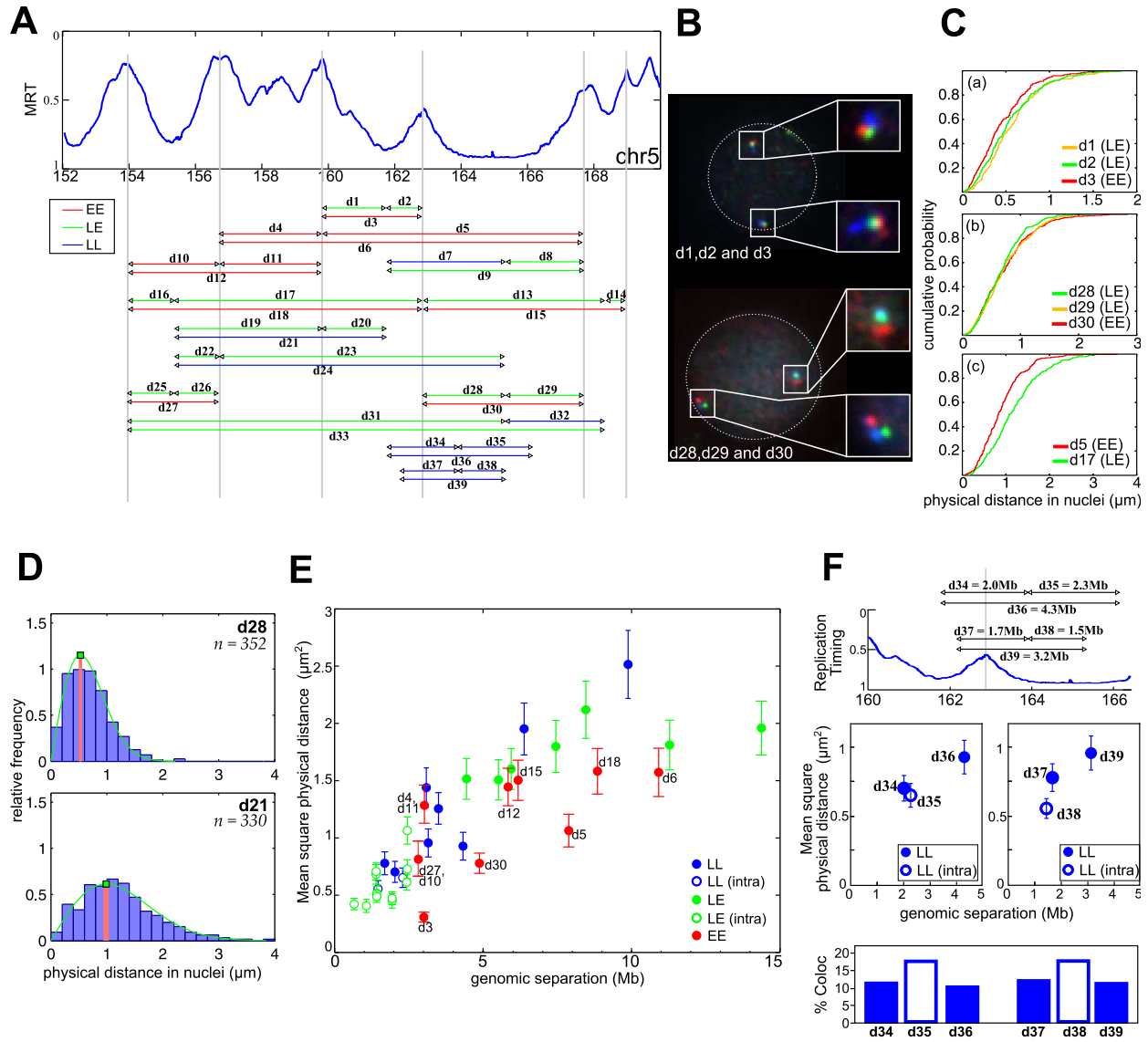


Figure 45: FISH distance measurements confirm the physical proximity between early replicated regions.

A- Physical distances measured by 3 colors FISH in resting PBMC. The arrowheads correspond to the positions of the used probes (table 4), and the line to the measured distances. Three sets of measurements were performed: EE between two timing peaks, LL between two timing valleys, and LE between a timing peak and a timing valley. **B-** Example of FISH images in resting PBMC nuclei (d1=GB, d2=BR, d3=GR, d28=BR, d29=GR, d30=GB). On the two chromosomes 5, the physical distances separating the gravity centers of the three spots were measured in more than one hundred PBMC nuclei. **C-** Cumulative distance measurements. (a) d3 (3Mb, red) is smaller than d1 (1.9Mb, yellow) and d2 (1.1Mb, green). (b) d28, d29, d30 have the same physical distance even if d30 (red) is 5Mb and d28 (green), d29 (yellow) are about 2.5Mb. (c) d5 (red) is smaller than d17 (green) even if they both cover 7.6Mb. **D-** Relative frequency histograms of two measured distances (d28, 2.5Mb, top and d21, 6.4Mb, bottom). The green lines correspond to the fit of the frequency histograms by a Rayleigh law (Methods) **E-** From the Rayleigh fit parameter σ , we obtained the mean squared physical distance [Yokota et al., 1995] that we plotted as a function of the genomic separation (mean squared physical distance \pm 95% confidence interval). Red dots: measurement between two timing-peaks (EE) ; blue: between two late regions (LL) ; green: between a peak and a timing valley (LE). Open dots correspond to distances measured inside a single replication domain; close dots correspond to distances measured between different replication domains. Here, a replication domain is defined by its two timing peak borders surrounding a late-replicated core. **F-** Physical distances and interaction frequencies between two probes localized within the same replication domain or between two adjacent replication domains. Top: corresponding distances. Middle: as in E, mean squared physical distances as a function of the genomic separation. Bottom: percentage of cases where measured distances are below $0.3 \mu\text{m}$ (criteria used for colocalization).

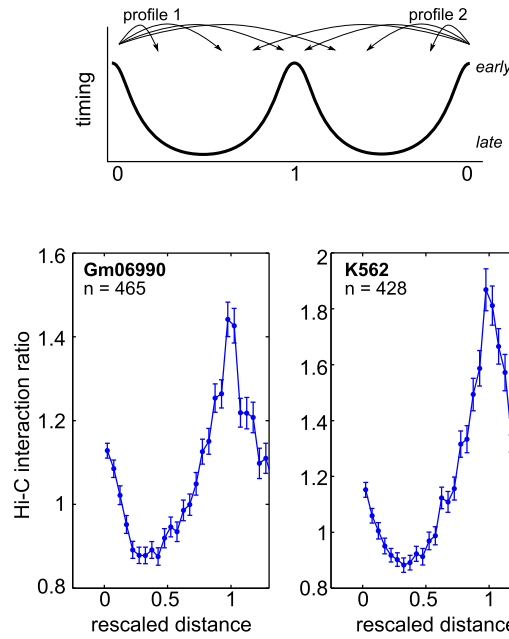


Figure 46: Hi-C interaction profile across timing peaks. Average Hi-C interaction ratio across timing peaks localized at the border between two consecutive U-domains (top, see Methods). Bottom, the mean interaction ratio profile as a function of the genomic distance rescaled so that U-domain size be 1. It corresponds to an average over $n=465$ (resp. $n=428$) timing peaks in Gm06990 (resp. K562) cell line.

replicated sequence with other regions and thus contributes to define independent megabase sized self-interacting structural domains in the nucleus.

4.3- Discussion

DNA replication exemplifies the difficulties to address the question of structure-function relationships in the nucleus. Indeed, some genomic and epigenetic features are of uttermost importance in the definition of replication origin (reviewed in Méchali [2010]) while replication itself leaves some evolutionary imprints in the genome [Brodie of Brodie et al., 2005; Touchon et al., 2005; Chen et al., 2011]. Here, we performed a series of 4C-Sequencing (4C-Seq) and FISH experiments that reveal to which extent the 3D-conformation of chromatin is linked to the replication timing.

At the local scale (1-2 Mb), that is the typical size of a replication timing domain [Hiratani et al., 2008; Ryba et al., 2010; Baker et al., 2012], we found that adjacent timing peaks are prone to physically interact, by looping out the intervening late-replicated sequences (figure 37 and figure 39). We recently proposed a segmentation of about half of the genome into U-domains (see Methods) defined by a late-replicating core surrounded by two timing peaks [Baker et al., 2012]. When averaging the Hi-C data over these U-domains, we recovered the preferential interactions between adjacent timing peaks in Gm06990 cells (figure 46, left), which corroborates genome-wide our conclusions. Moreover, we also noticed similar behavior in the chronic myelogenous leukemia K562 cell line (figure 46, right), indicating that preferential interactions between adjacent timing peaks

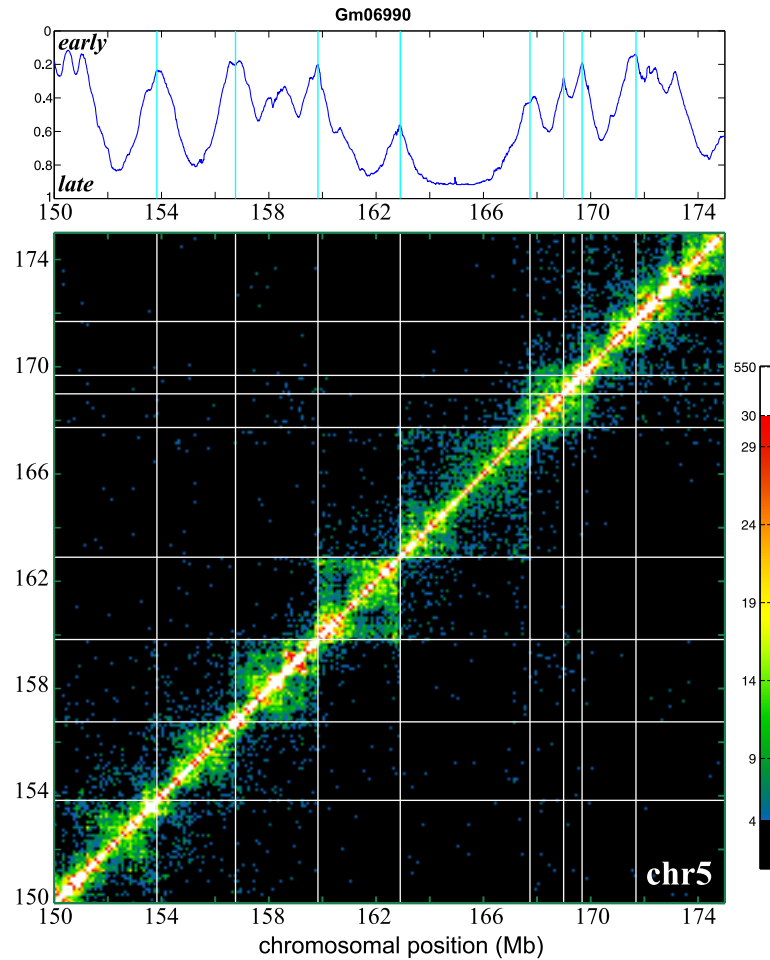


Figure 47: (suppl.) Replication domains correspond to structural self-interacting domains. Hi-C proximity heat-map for the intra-chromosomal interactions in Gm06990 cells on the chr5:150-175Mb locus. Each pixel represents the interaction in a 100 kb window. The colormap represents the number of reads for every pixel. A replication domain delimited by two local timing peaks corresponds to a self-interacting domain in Hi-C data, indicating that the timing peaks delineate the frontiers of megabase-sized structural domains.

are not limited to lymphoid cell types. Nevertheless, our 4C experiments allow to go much further in details since most of the local interactions (figure 37, dots) are not detectable in Hi-C matrices (figure 47). In addition to the preferential interaction between adjacent timing peaks, we also observed that frontiers limiting the DNA contacts of late replicating regions are found at timing peaks. This was observed by 4C (figure 37 and figure 39) and by FISH (figure 45F). These structural frontiers at timing peaks might rely on CTCF, which has been shown to be enriched at the boundaries of chromatin domains [Cuddapah et al., 2009]. Genome-wide analysis indeed revealed that CTCF is enriched in early-replicated regions [Ryba et al., 2010], and particularly at local timings peaks [Baker et al., 2012]. Since the timing peaks, enriched in CTCF binding sites, delineate boundaries of chromosomal domains, we propose to define a structural replication domain, not by its homogenous replication timing, but rather by its U-shape consisting in a late-replicated core surrounded by two local early-replicated peaks. Therefore, our definition is complementary to the definition of functional domains based on the replication program [Gilbert et al., 2010].

At the 10 Mb scale, we also showed preferential interactions between timing peaks (figure 37B, 39 and 45E). Similarly, late replicated viewpoints (V3, V4) preferentially interact with timing valleys. However, their interactions are spread over almost the entire length of the valley, indicating a relative lack of specific partners (figure 37B and 39). This might explain why we did not detect any physical proximity by FISH between late-replicated sequences (figure 45E).

At the scale of entire chromosomes, we demonstrated that chromosomal domains with an opposite replication timing are separated in the nuclear space (figure 41). Previous works that compared Hi-C data and replication timing values, already proposed that synchronously replicated sequences interact more frequently than expected [Yaffe et al., 2010; Ryba et al., 2010]. However, as compared to the Hi-C approach, the 4C protocol provides a much higher number of reads for each viewpoint, thereby allowing the direct detection of chromosome-wide interactions. This allowed us to show that the 4C signals of P1, P2 and P4 viewpoints faithfully reproduce the replication timing along the whole chromosome 5, which demonstrates the astonishing persistence, all over the chromosome, of a timing-linked long-range conformation of chromatin. Interestingly, it is not the precise value of the replication timing at a peak that determines long-range contacts but rather its relative value. Indeed, the mid-S-phase replicated P4 peak has a long-range interactions profiles similar to those of P1, P2 and P3 (figure 41A-B and figure 42).

This spatial segregation between early- and late-replicated domains is in agreement with previous results showing the relative absence of interactions between active (early-replicated) and inactive (late-replicated) chromatin [Simonis et al., 2006; Lieberman-Aiden et al., 2009; Sexton et al., 2012]. The nuclear processes explaining this separation remain to be identified. Insight may come from the P3 viewpoint that is localized in transcriptionally silent chromatin, about 100 kb away from the nearest expressed gene in PBMC. P3 exhibits in this cell type a positive correlation with the timing, albeit slightly lower than the one measured for P1, P2 and P4 viewpoints that are localized within expressed genes (figure 36 and 43). This might indicate that the influence of transcription and replication probably add up to enhance the long-range interaction capabilities of selected locus, maybe through the recruitment to spatially-coordinated replication and transcription factories [Malyavantham et al., 2008]. This hypothesis is further sustained by the results obtained for P5. This viewpoint lies in a weakly-expressed gene (figure 36) and is not located on a sharp timing peak (350kb-shoulder in the replication timing curve, (figure 37). In that context, we assume that the contribution of both transcription and replication are weaker, explaining why P5 interactions with timing peaks cannot be detected beyond the adjacent P4.

Our analysis revealed another striking property of chromatin organization in lymphocytes associated with replication. Indeed, a high similarity in local and long-range chromatin contacts was observed between resting and activated PBMC (figure 37C, 41C). This was unexpected since lymphocyte activation is associated with major changes in gene expression [Stentz and Kitabchi, 2004; Argypoulos et al., 2004; Hess et al., 2004; Chtanova et al., 2005; Grigoryev et al., 2009] and major histological modifications of the nucleus [Kyselá et al., 2005; Branco and Pombo, 2006]. In our experiment, PBMC cells were efficiently pushed out of dormancy since more than 40% of the PHA-stimulated cells are in S or G2/M phases (figure 38) and another fraction is probably also cycling through G1, albeit not detected in our analysis. The absence of differences between resting and activated lymphocytes therefore does not result from an inefficient PBMC activation by PHA. In addition, we also did not detect major changes between resting lymphocytes and lymphoblastoid

cells. This implies that (a) the histological disappearance of heterochromatin during lymphocytes activation takes place without disruption of the long-range interactions, and (b) the decondensation occurs at another scale than the one we are looking at through 4C. We can imagine that, by a dilatation of the nucleus and of the chromatin loops it contains, it should be possible to switch from the small compacted nucleus containing condensed chromatin loops to the larger nucleus of activated lymphocytes with decondensed chromatin loops, without altering the long-range contacts probed by the 4C approach. This transformation would maintain most of the neighboring relationships, maybe through constant anchoring at CTCF-enriched timing peaks.

At every scale analyzed, the 3D-conformation of chromatin is intimately linked to the underlying replication program. This 3D-conformation is remarkably conserved in the non-cycling, G0-blocked PBMC. The timing-linked conformation of chromatin described here is therefore not the resultant of the ongoing S-phase. This conformation may be an imprint of the previous S-phases accomplished during lymphocyte differentiation. [Courbet et al. \[2008\]](#) have already provided support for this idea by showing that the S-phase dynamics affects the anchoring of chromatin to the nuclear matrix in the subsequent cell cycle. Nevertheless, the cause to consequence relationship still has to be unraveled: instead of being conceived as an imprint of previous replication, chromatin conformation might result from some epigenetic memory driving the spatio-temporal program of the successive replication cycles.

4.4- Methods

Cell Culture Cells were grown at 37°C with 5% CO₂ in RPMI supplemented with 10% fetal bovine serum, glutamax and penicillin/streptomycin. PBMC, which are mostly lymphocytes [[Delves et al., 2006](#)], were isolated from healthy donors using Leucosep® (Greiner bio-one) and Ficoll-Paque PLUS (GE Healthcare) according to manufacturer's instructions. PBMC stimulation was performed by supplementing the culture medium with 1.5% Phytohemagglutinin (M Form, Life Technology) for 68 hr. A million stimulated PBMC were fixed with cold ethanol and stained with propidium iodide (20 µg/mL, 30 min 37°C) for cell cycle analysis (LSRII, FACSDiva™, BD Biosciences).

4C protocol and sequencing Cells were processed for 4C experiment as previously described [[Simonis et al., 2006](#); [Splinter et al., 2011](#); [Noordermeer et al., 2011b](#)] with only minor modifications. Cells were dispatched into 7 millions cell aliquots. All aliquots were fixed in PBS 10% SVF and 2% PFA (10 min, RT), lysed in 10 mM Tris, 10 mM NaCl, 0.3% NP-40 and dounced (B pestle). The cross-linked chromatin was digested overnight with 700 U of HindIII (Promega) and ligated with 150 U of T4 DNA ligase (Fermentas). The templates were further cut with 50 U Hin1II (Fermentas) and circularized with 300 U T4 DNA ligase (Fermentas). DNA was purified with phenol extraction followed by an ethanol precipitation, and chromatography (Roche High Pure PCR Product Purification Kit). DNA concentration was measured by fluorimetry (Quant-iT™ dsDNA Assay Kit, Broad Range, Life Technology). For every primer pair used, 10 PCRs were performed on 200 ng input DNA (for primer sequences, see table 3). For primer design,

Name	Chromosome 5 annealing sequence
P1_Hind	GAGGTTTCAGAGTTCAAGAAAGCTT
P1_Hin1II	GTACCTCGCTGGTGAAAACATG
P2_Hind	CCTCAGTCCTTTAAAGTGAAGCTT
P2_Hin1II	CAAGCACAAACCAAGAACATG
P3_Hind	ATCACTTTATGTAGCAGTGAAGCTT
P3_Hin1II	GCATTATTAAGTAAAGTTTCCACCATG
P4_Hind	AAGAGAAACGGATTGTTATGAAAGCTT
P4_Hin1II	TGGGTTACTATGTAAGGCCTCC
P5_Hind	GAGGAAAGAGAAGAATGAGTGCATG
P5_Hin1II	CCCCTGACCTTCCCAGAAGCTT
S_Hind	TTAACTTGGTGAATAAATCAGAAGCTT
S_Hin1II	TCAAATAAGGACTCTGGCAGTTG
V1_Hind	GGCAACTTTTAGTTCAATTCAAAGCTT
V1_Hin1II	GAACAATGTCCCATCCATTCTAG
V2_Hind	GCTGGGCTCAGGTGAAAGCTT
V2_Hin1II	CAATGGAGCCACTGATTACACATG
V3_Hind	TTTGAGGGGATTATCAAGAAAGCTT
V3_Hin1II	GGTACAGATGATATGAAAATCATTAAAC
V4_Hind	TTATAATAAAAGGCAGAAAGGAAGCTT
V4_Hin1II	CAATGACCTCTACTGCAATCATG

Table 3: (suppl.) Primers used for the 4C-Seq experiment. Similar strategy than in [Splinter et al. \[2011\]](#) was employed: we used primers containing in 5' the adaptor required for Solexa sequencing and in 3' the chromosome annealing sequence. Primers were page-purified before the PCRs. The annealing sequence is shown in this table.

we used a strategy similar to that described in [Splinter et al. \[2011\]](#) and [Noor-dermeer et al. \[2011b\]](#). PCR products were sequenced using a Solexa Genome Analyser IIX or Solexa Hi-Seq 2000 (76bp). Reads were first processed to identify the viewpoints and then mapped on the human genome (UCSC hg18 assembly) using Seqmap software [[Jiang and Wong, 2008](#)]. Only reads annealing on HindIII restriction sites were kept for subsequent analyses.

4C analyses The RAW signal consists in the number of reads per restriction site. For local analyses, a running mean was applied on the RAW signal with a window size of 20 restriction sites. To compare local 4C profiles between cell lines (figure 37C) the running mean signal was divided by the total number of reads corresponding to the chromosome 5. For chromosome-wide and genome-wide analyses, we defined a LRIF variable as the product of the percentage of restriction sites covered by reads in a 500Mb-sliding window (coverage) times the mean number of reads in this sliding window. The sliding step was 100kb. To compare the LRIF signal between cell lines, its value was divided by the total number of reads on the chromosome 5. Maxima from the LRIF signal were detected using an amplitude threshold based algorithm (Matlab, The Mathworks). All plots were generated using Matlab.

Heat-map and hierarchical clustering Heat-maps of pairwise distances between 4C peaks were generated by quantifying the percentage of cases where the measured distances are smaller than 250 kb. The value of this percentage was encoded using a green-red colormap. Matlab linkage function was employed to perform the hierarchical clustering.

FISH experiments We selected PBMC to perform FISH experiments, thus avoiding ambiguities due to cells in S or G2 phase exhibiting replication doublings. Resting PBMC were swollen in 75mM KCl (30min, 37°C), fixed in methanol 3:1 acetic acid, and then spread on microscope glass slides. FISH

hg18_Start	hg18_End	name
153,892,419	154,043,563	CTD-2568P10
155,277,598	155,458,580	RP11-484D20
156,685,117	156,887,519	RP11-152N9
159,722,795	159,901,802	RP11-135K8
161,669,550	161,831,097	RP11-587H15
162,035,265	162,156,312	CTD-3080G20
162,736,912	162,903,669	RP11-1012
163,714,926	163,854,439	RP11-1122J9
165,164,376	165,335,162	CTD-3159G23
165,994,779	166,159,481	RP11-669E20
167,614,694	167,782,307	RP11-277P14
168,278,178	168,401,684	CTD-3188O5
168,909,434	169,092,364	RP11-486H5

Table 4: (*suppl.*) **BAC clones used for the FISH experiments.** BAC Clones were purchased from Life Technology. Their specificity was tested by restriction analysis and hybridization localization on metaphase spreads before any distance measurement.

probes were generated by random priming on BAC clones (table 4) using amino-allyl dUTP. Probes were then coupled with amine-reactive alexa fluorophores (A488, A555 or A647, Life Technology). The specificity of BAC probes was tested by both restriction profile and hybridization on metaphase spreads. Hybridization of denatured probes (80°C, 5min) on denatured nuclei (73°C, 2min20) were performed overnight at 37°C (in 50% formamide, 2X SSC, 10% dextran, 10µg ss-DNA, 1-3µg human CotI). Slides were mounted with fluoromount-G (Southern BioTech) and DAPI was used as DNA counterstain. Images were collected using a wide-field Zeiss Axioimager (63X, Photometrics CoolSNAP HQ camera) and analyzed with Image J. The Rayleigh fits of measured distances $f(\frac{x}{\sigma^2})\exp(\frac{-x^2}{2\sigma^2})$ were performed using Matlab (σ corresponds to the shape parameter). The mean square physical distance is defined as $\langle r^2 \rangle = 2\sigma^2$, according to Yokota et al. [1995].

Mean replication timing profiles and replication U-domains We used the mean replication timing (MRT) profiles along the complete human genome derived by Baker et al. [2012] from Repli-Seq data [Hansen et al., 2010]. The signal analysis performed in Baker et al. [2012] led to the segmentation of about half of the genome into replication timing domains corresponding to U-shaped MRT profiles (late-replicating core surrounded by two early-replicating loci). MRT profiles and U-domains coordinates for Gm06990 and K562 were obtained directly from the authors.

Whole genome Hi-C chromatin conformation data analysis We used the spatial proximity maps of the human genome generated using Hi-C method [Lieberman-Aiden et al., 2009]. We downloaded 100 kb resolution maps for Gm06990 and K562 cell lines from the GEO web site (GSE18199 binned heatmaps): <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18199>. Interaction counts between loci on same chromosome were normalized by the mean intra-chromosomal interaction count averaged over all pairs of loci separated by the same genomic distance. To properly analyze the behavior of interaction frequencies across timing peaks, we selected well-defined timing peaks that are shared by two consecutive U-shaped replication timing domains [Baker et al., 2012]. We then extracted, from the Hi-C interaction ratio matrices, two 4C-like interaction profiles converging to the shared timing peak: the first one coming from the leftmost U-domain border, and the second one coming from the rightmost U-

domain border. All 4C-like interaction ratio profiles were finally averaged after rescaling the distances so that the U domain size was 1 (figure 46, top).

Acknowledgements:

We are very grateful to D.M. Gilbert, O. Hyrien, H. Julienne and C. Vailant for stimulating discussions. This work is supported by grants from the ANR (REFOPOL, ANR 10 BLAN 1615), CNRS PEPS 2009, ARC grant (n°ECL2010R01122), and EMBO short-term Fellowships (ASTF-381.00-2008). It benefited from the expertise of the high throughput sequencing platforms of IMAGIF (Centre de Recherche de Gif - www.imagif.cnrs.fr) and MGX-Montpellier GenomiX (<http://www.mgx.cnrs.fr/>). The authors also thank PLATIM (PLAteau Technique d'Imagerie et de Microscopie, IFR128, Lyon FRANCE).

Topologie de la chromatine : Extension de l'analyse à d'autres échelles, points de vue et types cellulaires

5.1	Conformation de la chromatine dans des cellules pro-érythrocytaires	148
5.1.1	Conformation de la chromatine à proximité des points de vue	148
5.1.2	Partenaires préférentiels à longue-distance	151
5.2	Organisation locale de la chromatine à timing constant	152
5.2.1	Détail de la région et PCR 4C-Seq	152
5.2.2	Analyse des profils 4C	153
5.3	Analyse des partenaires chromatinien d'un plateau précoce	156
5.4	Les loci précoces sur plusieurs Mb <i>trans</i>-interagissent	158
5.4.1	Matériel et méthodes : Analyse des <i>trans</i> -interactions dans le Hi-C	158
5.4.2	Proportion d'interactions en <i>trans</i> le long du chromosome	159
	<i>Le trans-index est globalement homogène le long d'un chromosome.</i>	159
	<i>Des variations du trans-index ($\pm 10\%$) sont néanmoins visibles</i>	161
5.4.3	Chromosomes préférentiellement contactés	164

Le chapitre précédent souligne l'existence de nombreux liens entre organisation tridimensionnelle de la chromatine et réplication. En effet, les loci répliqués en début et en fin de phase S occupent des territoires nucléaires distincts et interagissent peu ensemble. Plus localement, nous avons montré que les séquences répliquées en fin de phase S forment des domaines structuraux d'auto-interaction dont les frontières coïncident avec les séquences répliquées en début de phase S.

Ces résultats sont acquis pour les cellules lymphoblastoïdes et PBMC (cellules mononucléées du sang périphérique). Dans ce chapitre, je souhaite poursuivre la description de l'agencement 3D de la chromatine en continuant de souligner le lien avec le timing de réplication. Toutes les analyses seront basées sur la méthodologie 4C ou Hi-C. Nous aborderons les 4 points suivants.

1. Afin de préciser l'étroit lien entre le timing de réplication et la conformation de la chromatine, nous avons effectué un 4C dans les cellules K562 (issues d'une leucémie myéloïde à caractère érythrocytaire) dont le timing de réplication présente une zone d'initiation précoce supplémentaire dans la région chr5:150-170Mb. La conformation de la chromatine des cellules K562 et lymphoblastoïdes se révèle-t-elle différente au niveau de ce pic de timing additionnel présent dans les cellules K562 ?
2. Jusqu'à maintenant, nous avons souligné l'existence d'interactions fortes entre deux loci précoces consécutifs, probablement *via* la formation d'une boucle comprenant les loci tardifs intermédiaires. Néanmoins, certaines régions du génome présentent un timing de réplication à peu près constant

sur plusieurs centaines de kilobases (plateau précoce ou plateau tardif). Nous étudierons comment s'agence la chromatine au sein d'un plateau précoce, alors que le timing de réplication ne varie pas. Existe-t-il des partenaires préférentiels au sein du plateau précoce ou le repliement de la chromatine est-il plus aléatoire ?

3. A l'aide de ce point de vue situé au sein d'un plateau précoce (point 2), nous analyserons le patron d'interaction du plateau précoce à plus longue-distance. Possède-t-il les mêmes partenaires chromatinien qu'un pic de timing ?
4. Enfin, nous changerons d'échelle pour tenter d'identifier, parmi tous les chromosomes, les partenaires préférentiels de chaque Mb du génome (*trans*-interactions). Peut-on identifier, à partir des données Hi-C, des portions de chromosomes établissant localement plus de *trans*-interactions, et donc potentiellement localisées en surface du territoire chromosomique ? Quel est leur timing de réplication ?

5.1- La conformation de la chromatine de cellules pro-érythrocytaires et lymphoblastoïdes ne diffère que lorsque le timing de réplication diverge

Le timing de réplication est globalement conservé dans la région chr 5:150-170Mb entre les cellules K562 (issues d'une leucémie myéloïde chronique à caractère érythrocytaire) et les cellules lymphoblastoïdes. Cependant, un pic de timing supplémentaire est présent dans les cellules K562. Cette différence offre l'opportunité de tester dans quelle mesure une modification du timing de réplication est associée à un changement de la topologie de la chromatine.

Pour aborder cette question, un 4C a été effectué avec les points de vue P1, P4, V3 et V4. Seuls ces 4 couples ont été retenus car les pics ciblés par les amorces P2 et P3 sont légèrement décalés dans les cellules K562, et les points de vue P5, V1, V2, et S conduisaient, dans les cellules lymphoblastoïdes, à des profils 4C parfois difficiles à interpréter. Le protocole expérimental utilisé avec les cellules K562 est celui décrit chapitre 3.

5.1.1- Conformation de la chromatine à proximité des points de vue

Les profils 4C correspondant aux deux types cellulaires sont représentés figure 48. Ils sont localement extrêmement similaires (les profils 4C de P1, P4 et V3 sont quasiment superposables). On retrouve avec P1 et P4 des interactions préférentielles entre pics de timing consécutifs. On note également, pour V3, des frontières nettes limitant les capacités d'interactions au niveau des pics de timing voisins.

Une différence flagrante entre les profils 4C des cellules lymphoblastoïdes et K562 réside autour de 163.8 Mb (zone grisée, figure 48). Dans les cellules K562, cette région interagit moins avec V4, et plutôt avec P1 et P4. Cette zone correspond également au seul changement clair de timing de réplication dans la région

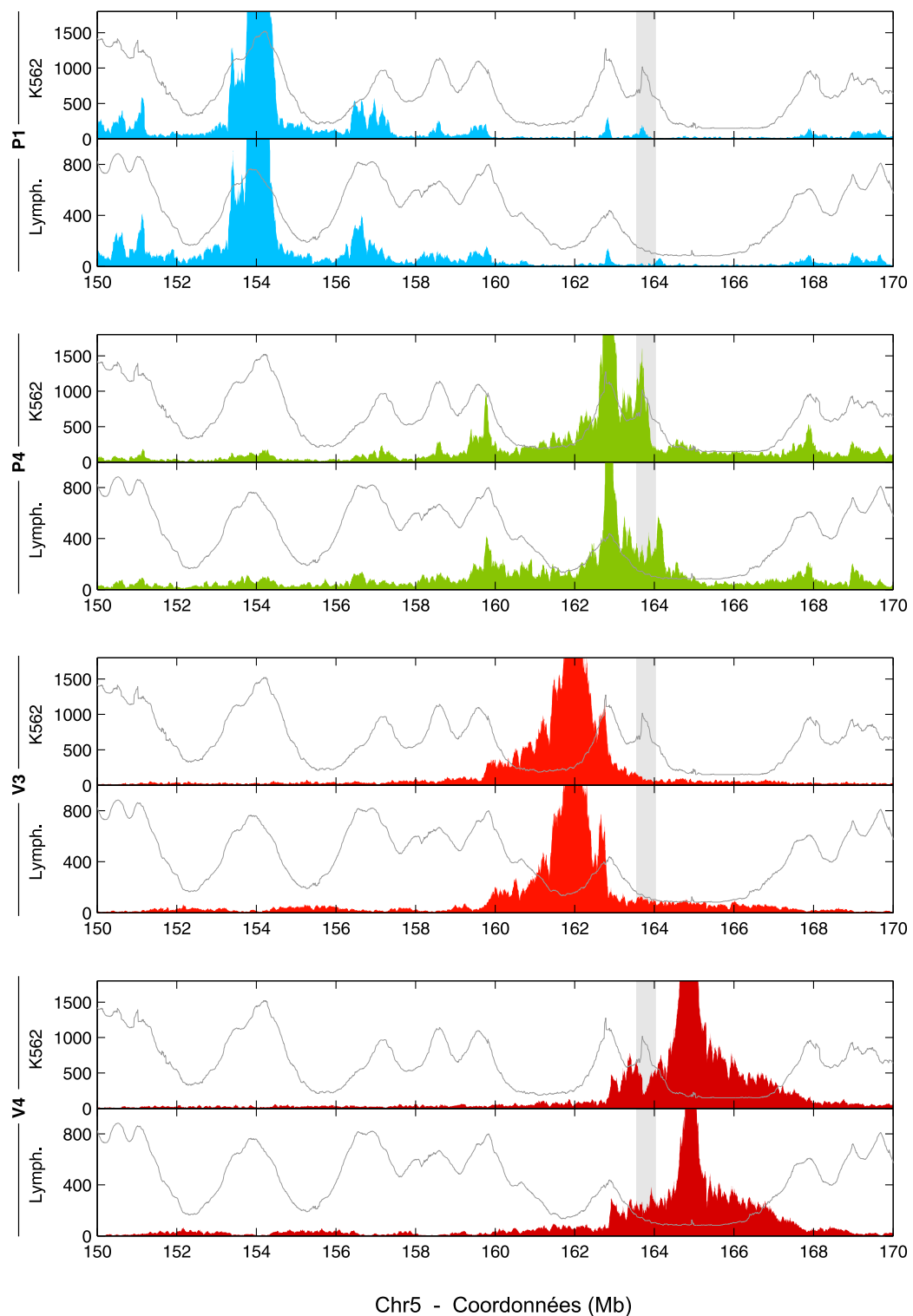


FIGURE 48 – Profils d’interactions locales des cellules lymphoblastoïdes et K562. Chaque groupe de panneaux correspond aux profils des fréquences d’interactions d’un point de vue (de haut en bas : P1, P4, V3 et V4 ; moyenne glissante sur 20 sites de restriction). Le panneau supérieur correspond aux cellules K562 et le panneau inférieur aux cellules lymphoblastoïdes. La courbe grisée de chaque panneau représente le timing de réplication moyen. La zone grisée (vers 163.8 Mb) marque un pic de timing présent uniquement dans les cellules K562.

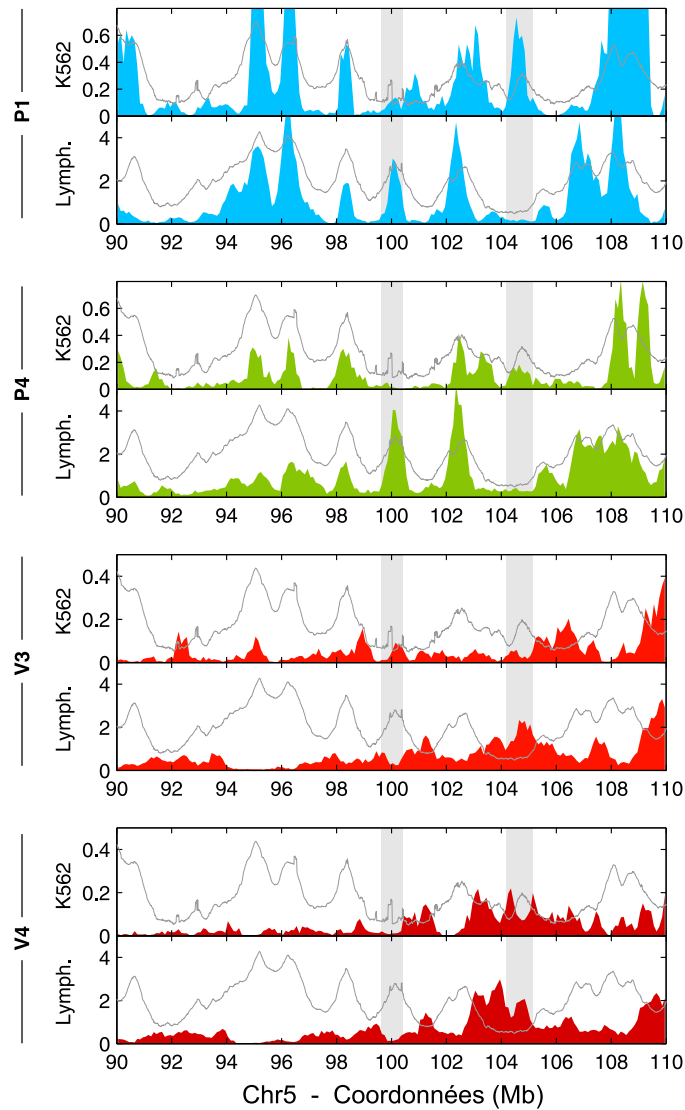


FIGURE 49 – Profils d’interactions à longue-distance des cellules lymphoblastoïdes et K562. Profil LRIF (Long-Range Interaction Frequency, voir méthodes au chapitre 4) pour les 4 points de vue. Les zones grisées (vers 100 Mb et 105 Mb) désignent des régions où le timing est différent dans les K562 et lymphoblastoïdes. La courbe grisée de chaque panneau représente le timing de réplication moyen.

chr5:150-170Mb. La présence de ce pic de timing additionnel dans les cellules K562 limite les capacités d'interactions de la région tardive voisine (ciblée par V4) et autorise les interactions avec les pics de timing à proximité (ciblés par P1 et P4). L'apparition de ce pic de timing pourrait être liée à l'expression d'un gène mal annoté autour de 163.8 Mb dont le produit a été retrouvé dans une banque de transcrits¹. Le promoteur potentiel de ce transcrit est marqué par H3K27ac et H3K4me3 dans les cellules K562, indiquant qu'il pourrait être exprimé.

En résumé, on retrouve dans les cellules K562 une organisation locale de la chromatine associée au timing de réplication. De manière remarquable, la seule différence flagrante de la topologie de la chromatine entre les cellules K562 et lymphoblastoïdes réside au niveau du pic de timing additionnel. Ce résultat renforce l'étroit lien entre le timing de réplication et la conformation de la chromatine.

La seule propriété qui n'est pas aussi visible dans les cellules K562 est l'interaction entre régions tardives voisines (des partenaires clairs de V3 et V4 sont observés autour de 152 Mb et autour de 155 Mb dans les cellules lymphoblastoïdes, mais pas dans les cellules K562, figure 48).

5.1.2- Partenaires préférentiels à longue-distance

Comme pour les cellules lymphoblastoïdes et les PBMC, les profils d'interactions à longue-distance des cellules K562 (LRIF, voir méthodes au chapitre 4) montrent que les régions précoces et tardives possèdent des partenaires différents. Les points de vue précoces (P1 et P4) interagissent majoritairement avec des régions correspondant à des pics de timing, tandis que les points de vue tardifs (V3 et V4) interagissent avec d'autres séquences tardives (figure 49). Régions précoces et tardives ségrègent donc dans l'espace nucléaire.

A deux reprises, sur la portion représentée figure 49, le timing de réplication diffère d'un type cellulaire à l'autre. Dans les cellules lymphoblastoïdes, un pic de timing est présent vers 100 Mb et, comme attendu, il s'agit d'un partenaire de P1 et P4. Dans les cellules K562, n'étant plus un pic de timing, cette région n'est plus contactée par P1 ou P4. Inversement, la région vers 105 Mb, qui correspond à un pic de timing dans les cellules K526, est un partenaire de P1 et P4. Les cellules lymphoblastoïdes montrent que cette région, qui est alors une vallée tardive, est plutôt contactée par les points de vue V3 et V4.

Pour résumé, on retrouve dans les cellules K562 la ségrégation des régions précoces et tardives dans l'espace nucléaire. Les changements de timing de réplication s'accompagnent de modifications des partenaires.

L'ensemble des liens entre timing de réplication et organisation de la chromatine est retrouvé dans les cellules K562. Aussi bien à quelques mégabases, qu'à quelques dizaines de mégabases, un changement de timing de réplication est associé à une modification de la topologie de la chromatine.

1. gène BC011998, sans cadre ouvert de lecture, d'après le site genome.ucsc.edu

5.2- Lorsque le timing de réplication est homogène sur 1 Mb, les contacts s'effectuent entre portions de chromatine partageant le même contexte chromatinien

Nous l'avons vu, des interactions observées de manière robuste existent entre deux pics de timing séparés par une région tardive. Les oscillations du timing de réplication corrélient avec des oscillations du profil d'interactions. Néanmoins, certaines portions du génome possèdent un timing de réplication à peu près constant sur plusieurs centaines de kilobases. Notamment, certaines d'entre-elles, que l'on nomme "plateau précoce" ont un timing de réplication proche de zéro (répliquées en début de phase S) homogène sur 1 Mb. Si les variations du timing de réplication contraignent l'organisation de la chromatine, comment cette dernière s'agence-t-elle lorsque le timing de réplication est à peu près constant sur 1 Mb ? Existe-t-il une topologie particulière de la chromatine alors qu'il n'y a pas de variations de grande ampleur du timing de réplication ?

Pour répondre à ces questions, un 4C a été effectué à l'aide d'amorces localisées 1.5 kb en amont de c-myc, en plein centre d'un plateau précoce. Ce point de vue est placé au niveau de l'ORI en amont de c-myc, région correspondant également à son promoteur. Précisons dès maintenant que ce 4C a été effectué avant ceux présentés au chapitre 4. La librairie 4C n'était alors peut-être pas d'aussi bonne qualité.

5.2.1- Détail de la région et PCR 4C-Seq

Cette étude a été réalisée sur des cellules lymphoblastoïdes. Deux couples d'amorces ciblant le même site HindIII ont été choisis : l'un, appelé MycL, est localisé sur le segment Hin1II-HindIII en amont du site HindIII ciblé, tandis que le couple MycR est localisé en aval (figure 50A).

Le couple MycR conduit à une amplification 4C qui rappelle celles observées avec les amorces du chr 5 (figure 50B) : (1) le patron de migration est sous-forme d'un "smear" avec une bande majoritaire de 400 bp correspondant au cas où le site HindIII n'est pas digéré, et (2) l'ensemble du "smear" est décalé lorsque les amorces contenant les adaptateurs de séquençage sont utilisées. Les amorces MycL conduisent à une amplification nettement moins robuste : pas de smear mais de nombreuses bandes, et des amplifications surviennent sur l'ADN génomique avec les amorces sans adaptateurs. Bien que les PCR 4C effectuées avec les amorces MycL ne satisfassent pas complètement les critères de qualité que nous avons établis plus tard au cours de l'avancée de ces travaux (chapitre 3), elles ont tout de même été séquencées.

Comme pour le locus du chr 5, la même PCR est répétée 10 fois et les produits d'amplification sont séquencés sur Solexa Genome Analyzer Iix. 99,8% des lectures sont attribuées à leur point de vue respectif. Pour les lectures correspondant aux amorces MycR, 80% des lectures sont efficacement alignées sur le génome humain et en très grande majorité exactement sur un site HindIII. En outre, 86% des lectures alignées sont localisées sur le chr 8, ce qui est en accord avec l'existence de territoires chromosomiques individualisés.

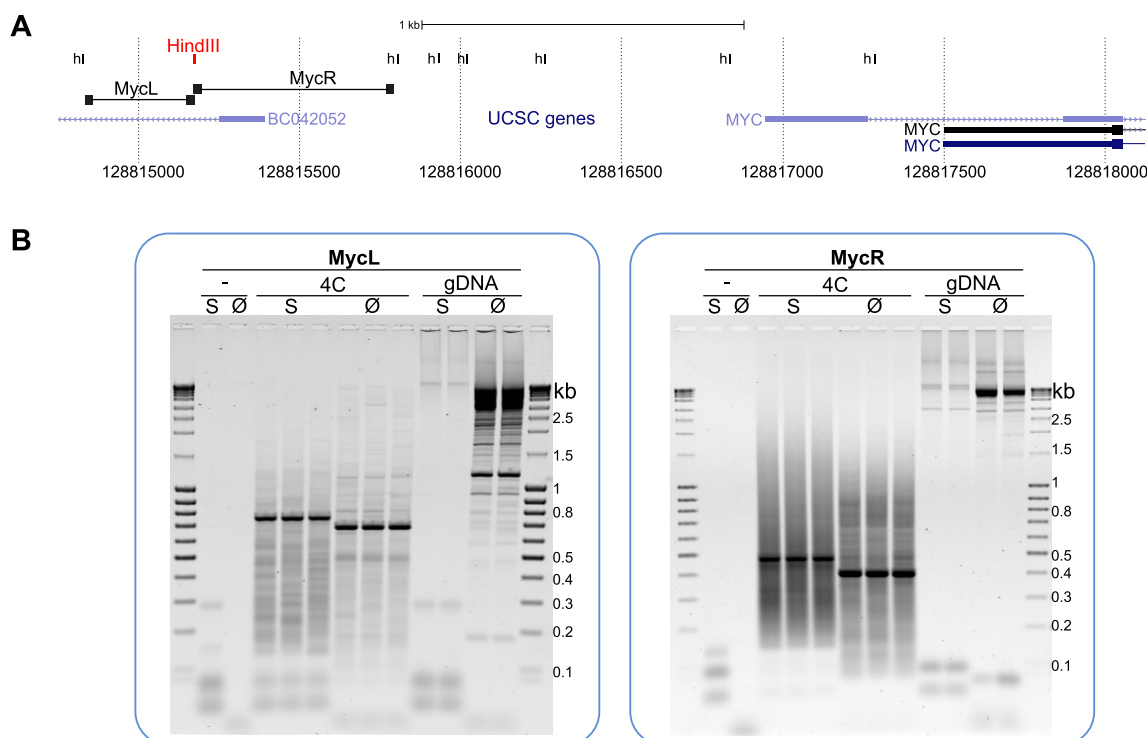


FIGURE 50 – PCR 4C avec comme point de vue l’origine de réplication à proximité de c-myc. **A-** Position des gènes et du site HindIII ciblé (UCSC genome browser, hg18). Deux couples d’amorces (MycR et MycL) ciblent le même site de restriction. **B-** Comparaison des profils de migration des PCR 4C effectuées avec des amorces contenant (S) ou ne contenant pas (Ø) les adaptateurs de séquençage. Les amorces MycL ne respectent pas tous les critères de qualités établis au cours de l’avancée de ces travaux et énoncés au chapitre 3.

5.2.2- Analyse des profils 4C

Les profils 4C correspondant sont représentés figure 51. Afin d’en faciliter la lecture, les zones décrites dans le texte sont appelées à l’aide d’une lettre minuscule encadrée.

Les profils obtenus avec MycL et MycR sont très reproductibles. Les profils obtenus avec les amorces MycR et MycL sont extrêmement similaires (figure 51, haut). Ceci témoigne de la reproductibilité du 4C en ciblant le même site HindIII par deux fragments de restriction différents (MycL et MycR, figure 50A).

Les partenaires préférentiels partagent les mêmes marques épigénétiques que le point de vue. Le point de vue 4C, tout comme de nombreux partenaires de celui-ci situés au sein du plateau précoce (128.3 Mb -**a**-, 128.4 Mb -**b**- et 128.6 Mb -**c**-) résident dans une chromatine enrichie en marques d’histones actives (H3K9ac, H3K27ac, H3K4me3, H3K36me3), en ARN-polymérase II (y compris phosphorylée), et en TBP.

Au demeurant, cette propriété reste vraie pour les partenaires situés en dehors du plateau précoce, par exemple à 130.6 Mb -**d**-, et dans une moindre mesure autour de 129.2 Mb -**e**- (toutes les marques énoncées avant, sauf TBP) et à 129.4 Mb -**f**- (toutes sauf TBP et Pol II). Notons d’ailleurs que des partenaires moins fréquents (126.25, 126.4, 126.7 Mb -**g**-), au niveau du pic de timing voisin) partagent, eux-aussi, la plupart de ces marques.

CHAPITRE 5. Topologie de la chromatine : extension de l'analyse à d'autres échelles, points de vue et types cellulaires

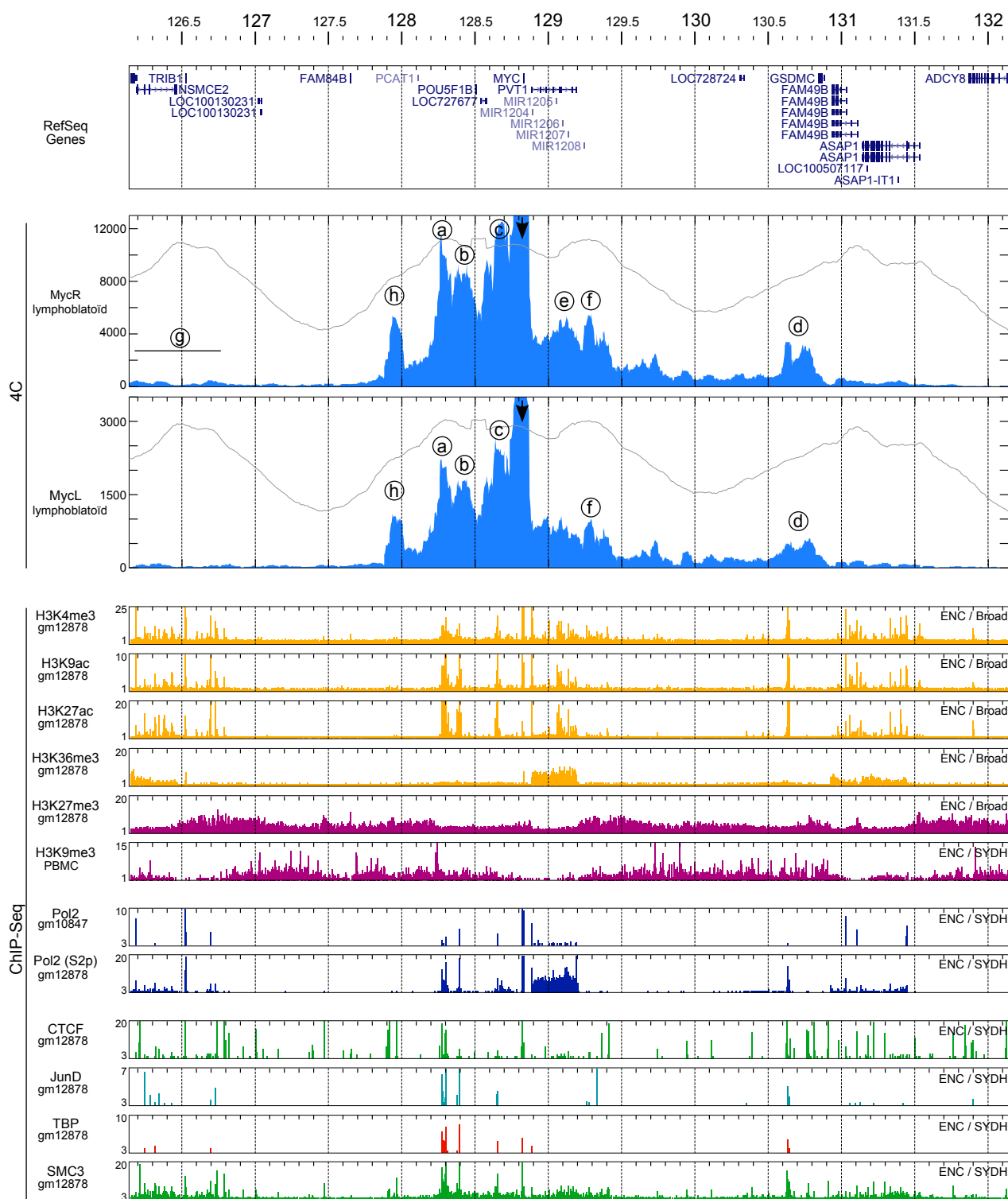


FIGURE 51 – Conformation de la chromatine au sein d'un plateau précoce. De haut en bas, les gènes (RefSeq), les profils 4C (moyenne glissante) avec le timing de réplication (gris), 6 marques d'histones (ChIP-Seq), la présence d'ARN-polymérase II et de l'ARN-pol II phosphorylée, CTCF, JunD, TBP (TATA-binding protein), et Smc3 (cohésines). Les panneaux de ChIP-seq représentent la moyenne du signal dans des lymphoblastoïdes (gm12878 et gm10847) sauf pour le panneau H3K9me3 où il s'agit de celui établi dans des PBMC (le profil H3K9me3 disponible pour les gm12878 est étrange car il corrèle avec les régions ouvertes et surtout il est complètement différent de celui obtenu pour les PBMC, Dnd41 (Leucémie-T) et K562 qui ont toutes à peu près le même profil). Ces données sont mises à disposition par le Broad Institute (ENC / Broad) ou par Stanford/Yale/USC/Harvard (ENC / SYDH). Les coordonnées correspondent à hg18 (chr 8). Le décalage entre les coordonnées hg18 (4C) et hg19 (ChIP-Seq, RefSeq) a été pris en compte dans l'alignement des panneaux. Les lettres cerclées font référence au texte ; la flèche indique la position du point de vue.

Les partenaires préférentiels correspondent à des sites de fixation de facteurs de transcription. Les partenaires de c-myc au sein du plateau précoce (128.3 Mb -**(a)**-, 128.4 Mb -**(b)**-, 128.6 Mb -**(c)**- et 129.4 Mb -**(f)**-) correspondent à des sites de fixation de facteurs de transcription. Le ChIP-Seq de JunD est donné comme exemple (figure 51), mais d'autres facteurs de transcription sont également observés sur ces sites, comme IRF3 (Interferon Regulatory Factor 3), EBF1 (Early B-cell Factor 1), STAT1 ou STAT3 (non-exhaustif et non-illustré).

Ceci reste vrai pour la plupart des partenaires situés en dehors du plateau précoce (130.6 Mb -**(d)**- et autour de 126.5 Mb -**(g)**-).

Une signature partagée par tous les partenaires est la présence de CTCF ou des complexes cohésines. Tous les contacts locaux établis au sein du plateau précoce s'effectuent avec des loci comportant des sites de fixation pour CTCF et pour les complexes cohésines (**(a)**, **(b)**, **(c)**, **(e)** et **(f)**).

Ceci reste valable pour les partenaires situés en dehors du plateau précoce (**(g)**, **(d)** et **(h)**).

Un seul partenaire ne possède aucune marque de chromatine active évoquée ci-avant. Seul le partenaire situé peu avant 128 Mb -**(h)**- ne possède aucune marque de chromatine active évoquée précédemment. En effet, cette région n'est pas spécialement ouverte et sans polymérase II. Néanmoins, on note la présence de CTCF, des complexes cohésines, et de certains facteurs de transcription (EBF1 et STAT1 par exemple).

Ne s'agissant pas d'un pic de timing (même si une inflexion dans le timing de réplication est visible à ce niveau), le processus nucléaire sur lequel repose cette interaction reste une énigme.

Des contacts privilégiés s'établissent au sein d'une grande région à timing homogène. En utilisant un point de vue situé au cœur d'un plateau précoce, nous avons détecté des partenaires chromatinien dans la zone où le timing de réplication reste à peu près constant. Ces partenaires partagent avec le point de vue les mêmes marques épigénétiques et la même machinerie d'expression (facteurs de transcription, polymérase). En absence de variations du timing de réplication, la chromatine est donc organisée de telle sorte à favoriser les contacts entre séquences partageant le même contexte chromatinien. On retrouve donc, à cette échelle très locale, une ségrégation en fonction du niveau d'ouverture de la chromatine : les loci les plus ouverts du plateau précoce (qui correspond déjà, dans l'ensemble, à de la chromatine ouverte) et les loci un peu moins ouverts interagissent peu ensemble. Au sein d'un plateau précoce, le point de vue choisi ne contacte donc pas avec la même probabilité tous les loci précoces.

Si les oscillations du timing de réplication s'accompagnent d'oscillations dans les profils d'interactions (chapitre 4 et section 5.1), des contacts préférentiels s'établissent également lorsque le timing de réplication est à peu près constant. Les contacts observés au sein d'un plateau précoce s'effectuent entre loci partageant les mêmes marques épigénétiques, la même machinerie d'expression (facteurs de transcription, polymérase), et présentant des sites de fixation pour CTCF et les complexes cohésines.

5.3- Comme les pics de timing, un plateau précoce interagit avec d'autres séquences répliquées en début de phase S

Le 4C effectué dans les cellules lymphoblastoïdes à l'aide d'amorces en amont de c-myc nous permet également d'identifier des partenaires situés au-delà du plateau précoce. Correspondent-ils à d'autres loci répliqués en début de phase S ? Les partenaires d'un plateau précoce, à plusieurs mégabases et à plusieurs dizaines de mégabases, sont-ils similaires à ceux que pourrait contacter un pic de timing ?

Pour répondre à ces questions, nous avons ajusté l'axe des ordonnées des profils 4C afin d'étudier ces interactions plus lointaines. La démarche employée est similaire à celle exposée sur le chromosome 5.

Passé le plateau précoce, les profils 4C sont moins reproductibles. Bien que ciblant le même site HindIII, les profils 4C de MycL et MycR apparaissent, à cette échelle-ci, moins reproductibles (comparer la figure 51 et 52A). Ceci pourrait être dû à un plus faible taux d'échantillonnage. En effet, si le pourcentage de sites HindIII capturés est supérieur à 40% (dans une fenêtre de 500 kb) entre 128 et 131 Mb, il chute en dessous de 15% (MycR) ou 10% (MycL) pour les loci situés à plus de 4 Mb du point de vue (à titre de comparaison, sur le chr 5, le

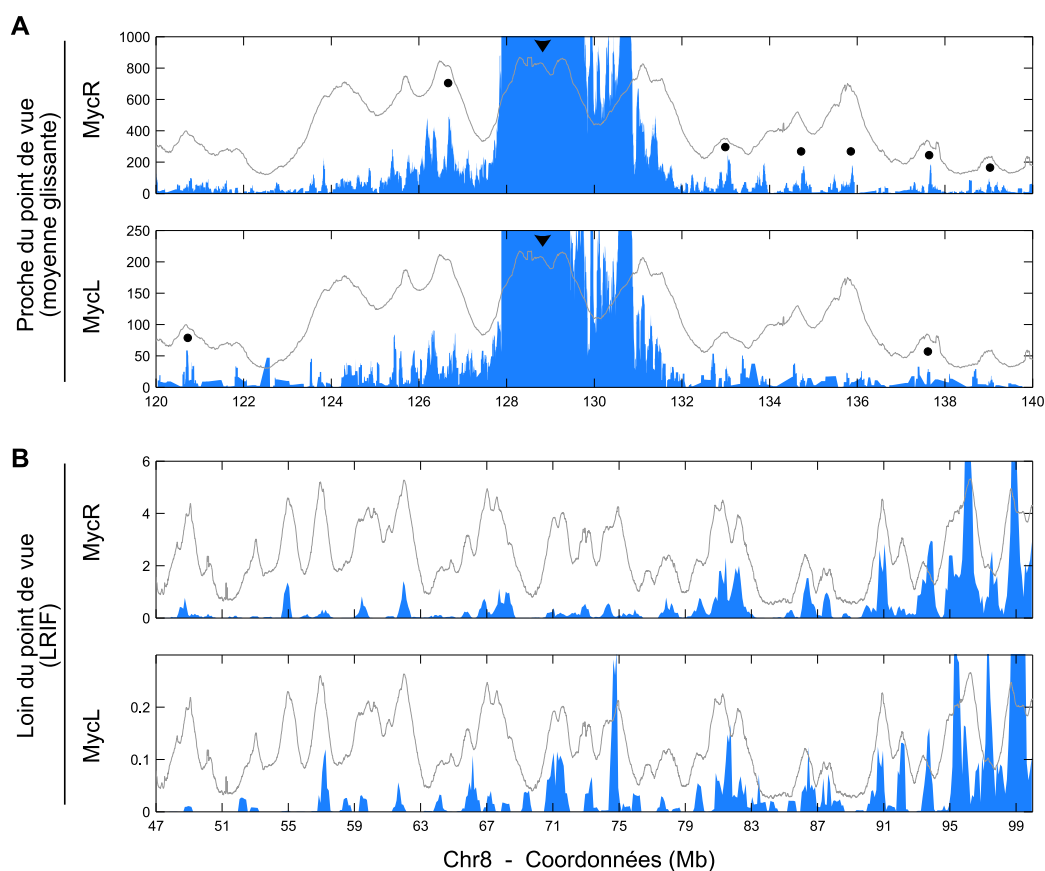


FIGURE 52 – Un plateau précoce interagit préférentiellement avec d'autres loci précoces. A- Profil 4C local (moyenne glissante sur 20 sites de restriction) ; la tête de flèche indique la position du point de vue, les points noirs indiquent des rebonds au niveau de pics de timing. B- Interactions à longue-distance (LRIF, voir chapitre 4). En A et B, le timing de réplication moyen est représenté par une courbe grise.

pourcentage de site HindIII capturés est toujours supérieur à 50% pour des séparations génomiques similaires). Ce faible pourcentage augmente le risque de rater certaines interactions (faux-négatifs), mais aussi de détecter de fausses interactions (le rapport signal sur bruit étant bas, la capture d'un site de restriction isolé et amplifié par PCR contribuera beaucoup à la valeur de la moyenne glissante). Une librairie 4C de moindre qualité et/ou une amplification PCR moins efficace pourraient être à l'origine ce sous-échantillonnage. Cette seconde possibilité s'applique au moins à MycL puisque la PCR correspondante est peu robuste (figure 50B). Ceci confirme que la sélection des amorces est particulièrement critique pour effectuer un 4C de qualité.

Néanmoins, il est également possible que cette faible reproductibilité soit le reflet de l'impossibilité d'établir des interactions au-delà du plateau précoce. Les règles déterminant les interactions locales d'un plateau précoce pourraient être significativement différentes de celles des pics de timing, de sorte que les interactions soient très majoritairement limitées à la zone précoce entourant du point de vue.

Certains partenaires locaux correspondent à des pics de timing. Même si les profils 4C sont moins reproductibles, certains rebonds ponctuels (points noirs) coïncident avec des pics de timing (figure 52A). Le plateau précoce interagit donc avec certains pics de timing voisins. Néanmoins, la correspondance entre ces rebonds et les pics de timing est nettement moins forte que celle observée sur le chromosome 5 (figure 41A) car tous les pics de timing ne sont pas contactés et car certains partenaires ne correspondent pas à des pics de timing. Ceci indique que le plateau précoce, même s'il peut contacter d'autres pics de timing, a dans l'ensemble un comportement différent des pics de timing tel que P1, P2, P3 P4 du chr 5.

Gardons néanmoins à l'esprit que cette différence de comportement pourrait aussi être le reflet des biais d'échantillonnage (paragraphe précédent).

Les partenaires à longue-distance correspondent également à des pics de timing. Pour réduire les problèmes de sous-échantillonnage qui peuvent artificiellement changer la moyenne locale, nous avons défini dans le chapitre 4 une variable appelée LRIF qui renforce l'impact des lectures présentant une distribution dense. Appliquée au 4C sur ce locus, les profils obtenus à plusieurs dizaines de Mb du point de vue montrent des oscillations dont les maxima sont le plus souvent observés au niveau de régions répliquées en début de phase S (figure 52B). A longue distance, le point de vue situé au centre d'un plateau précoce interagit donc préférentiellement avec d'autres régions répliquées en début de phase S. On peut donc conclure que les interactions à longue-distance d'un plateau précoce sont similaires à celles établies par les pics de timing.

Un plateau précoce est capable d'interagir à longue-distance avec d'autres loci précoces. On retrouve donc le régime d'interactions à longue-distance des pics de timing (chromosome 5). Le régime d'interactions locales est nettement moins bien observé, puisqu'un plateau précoce, à la différence des pics de timing, n'interagit pas avec tous les pics de timing voisins. Ceci suggère que les règles déterminant les interactions locales sont donc sensiblement différentes pour un pic de timing et un plateau précoce.

5.4- Conformation de la chromatine au sein du territoire chromosomique : les régions précoces sur plusieurs Mb *trans*-interagissent ensemble

L'observation par hybridation *in-situ* de chromosomes entiers a montré que la chromatine de chaque chromosome est confinée dans un territoire nucléaire discret. Ces territoires sont relativement individualisés et agencés radialement de telle sorte que les petits chromosomes riches en gènes et à forte teneur en GC adoptent un positionnement central et que les grands chromosomes, moins denses en gènes, se positionnent en périphérie du noyau (page 33 et suivantes).

Comment la chromatine s'organise-t-elle au sein du territoire chromosomique (CT)? Certaines séquences sont-elles préférentiellement retrouvées en surface du CT, tandis que d'autres sont localisées au cœur du CT? Les séquences en surface du CT possèdent-elles des chromosomes partenaires particuliers? Les séquences répliquées en début de phase S sont-elles plutôt en surface du CT? Sont-elles des *trans*-partenaires préférentiels?

Certaines séquences constitutivement actives (comme la région 11p15.5 chez l'homme [Mahy et al., 2002]) ou activées lors du développement (comme le locus Hoxb, Hoxd chez la souris [Chambeyron and Bickmore, 2004; Morey et al., 2007]) sont fréquemment extrudées de leur territoire chromosomique ou localisées en périphérie de celui-ci. Cette localisation en surface ou en dehors du CT leur permet d'établir des contacts avec d'autres chromosomes (*trans*-contacts) qui peuvent être détectés par des approches basées sur les méthodologies de type "Chromosome Conformation Capture" [Würtele and Chartrand, 2006].

Pour étudier comment la chromatine se replie au sein du territoire chromosomique et comment se répartissent les régions précoces dans ce repliement, nous avons défini, le long du chromosome, un index quantifiant la capacité à établir des *trans*-contacts. Un index élevé devrait attester d'une localisation en périphérie ou en dehors du territoire chromosomique. Pour déterminer ce *trans*-index nous avons utilisé les données Hi-C de Lieberman-Aiden et al. [2009] qui fournissent un jeu d'interactions 2 à 2, certaines correspondant à des contacts au sein d'un même chromosome (*cis*) et les autres à des contacts entre chromosomes (*trans*).

5.4.1- Matériel et méthodes : Analyse des *trans*-interactions dans le Hi-C

Proportion d'interactions en *trans* Cette étude est réalisée à l'aide des matrices Hi-C à 1 Mb (contacts observés) téléchargées à partir du site GEO (GSE18199 binned heatmaps, <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18199>). Le long de chaque chromosome (pas = 1 Mb), nous avons dénombré le nombre de ligations en *cis* (N_{cis}^{Total} , au sein d'un même chromosome) et nombre de ligations en *trans* (N_{trans}^{Total} , entre deux chromosomes différents). A partir de ces deux valeurs pour chaque Mb de chaque chromosome, nous avons calculé un *trans*-index traduisant la capacité à établir des *trans*-interactions :

$$trans\text{-index} = \frac{N_{trans}^{Total}}{N_{cis}^{Total} + N_{trans}^{Total}} \cdot$$

5.4- Les loci précoces sur plusieurs Mb *trans*-interagissent

Cet index représente pour chaque Mb la proportion d'interactions s'effectuant en *trans*.

Interactions en *trans* avec un chromosome donné Le long de chaque chromosome, nous avons également calculé la proportion de *trans*-interactions impliquant un chromosome particulier. Ceci nous a permis de déterminer un *trans*-partenaire-index définit pour chaque Mb comme :

$$\text{trans-partenaire-index} = \frac{N_{trans}^{chrA}}{(N_{cis}^{Total} + N_{trans}^{Total}) \times \text{Taille séquencée chrA}} ,$$

où N_{trans}^{chrA} correspond au nombre de ligations en *trans* impliquant le chr A et N_{trans}^{Total} au nombre total de ligations en *trans*. Cet index a été calculé pour chaque Mb du génome et traduit la capacité de celui-ci à contacter le chrA (ou A = 1,2,...,X). Un grand chromosome occupant une fraction plus grande du noyau, il aura plus de chance d'être contacté qu'un petit chromosome (voir plus loin). Pour pouvoir comparer les valeurs du *trans*-partenaire-index, la taille séquencée du chromosome contacté a été prise en compte.

Corrélations épigénétiques Le *trans*-index et *trans*-partenaire-index ont été comparés à des marques épigénétiques obtenues à partir du site UCSC (hg19 : wgEncodeBroadHistoneGm12878H3k27me3StdSig, wgEncodeBroadHistoneGm12878H3k27me3StdSig, wgEncodeBroadHistoneGm12878H3k36me3StdSig, wgEncodeSydhHistonePbmCh3k9me3UcdSig). Celles-ci ont été ré-échantillonnées à 200 kb et rétro-converties en coordonnées hg18 (outil LiftOver) qui est l'assemblage du génome utilisé pour les matrices Hi-C. Tous ces profils de marques épigénétiques ont été établis dans les cellules lymphoblastoïdes Gm12878, proches des cellules Gm06990 utilisées dans le Hi-C, sauf pour le profil H3K9me3 qui a été établi dans des PBMC (raison évoquée dans la légende de la figure 51).

5.4.2- Proportion d'interactions en *trans* le long du chromosome

Pour tenter d'identifier des portions de chromosomes interagissant plus souvent en *trans* que leurs voisines, le *trans*-index (voir méthode) a été calculé pour chaque mégabase de chaque chromosome. Il est représenté le long de trois chromosomes (1, 12 et 21) sur la figure 53.

a- Le *trans*-index est globalement homogène le long d'un chromosome.

En première approximation, le *trans*-index est globalement homogène le long d'un chromosome puisque l'amplitude maximale de ses oscillations est de $\pm 10\%$. Chaque Mb d'un chromosome établit donc en proportion autant d'interactions en *trans*. Ainsi, $\sim 55\%$ de toutes les interactions détectées pour chaque Mb du chr 1 impliquent un autre chromosome (médiane sur les 230 Mb séquencés du chr 1); ce pourcentage atteint une médiane de 59% sur le chr 12 et 65% sur le chr 21 (figure 53, haut).

Pour ces trois chromosomes, la valeur médiane du *trans*-index est d'autant plus élevée que le chromosome est petit (55% pour le chr 1, 59% pour le chr 12

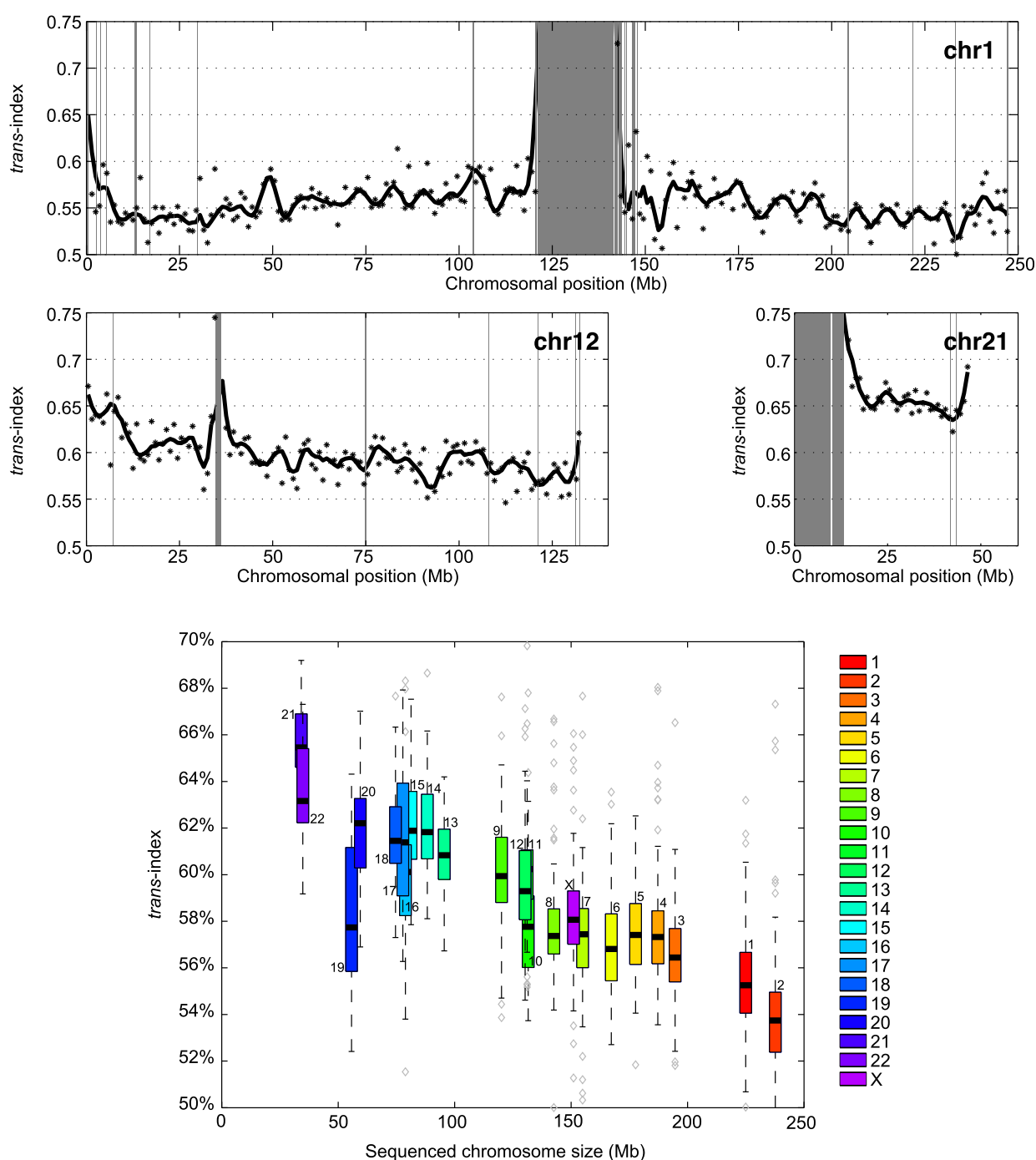


FIGURE 53 – Proportion d'interactions en *trans* : variation d'un chromosome à l'autre.
Haut-Le *trans*-index (voir méthode) est représenté pour chaque Mb des chromosomes 1, 12 et 21 (astérisque). Le signal est lissé par une régression polynomiale locale excluant les points situés à plus de 6 écarts-types (courbe pleine). Les zones non-séquencées (table ucsc GAP : centromères, télomères, absences de recouvrement entre contig, clones...) sont représentées par des rectangles grisés.
Bas- Pour chaque chromosome, le *trans*-index est représenté par un diagramme en boîte à moustaches. L'abscisse de chaque boîte correspond à la taille séquencée (GAP exclu) du chromosome. En les classant par leur taille séquencée, l'ordre des chromosomes est légèrement changé : le chr 2 apparaît plus grand que le chr 1, le chr 9 est plus petit que le chr 12, et le chr 20 plus grand que le chr 19.

et 65% pour le chr 21). Cette tendance est confirmée grâce à la représentation du *trans*-index de chaque chromosome en fonction de la taille séquencée du chromosome (figure 53, bas). Les grands chromosomes interagissent plus avec eux-mêmes que ne le font les petits chromosomes. Ce résultat pourrait s'expliquer par les volumes relatifs occupés par chaque chromosome dans le noyau. Un grand chromosome comme le chr 1 représente à lui seul près de 7,5% du génome total (et *a priori* 7,5% du volume nucléaire), alors qu'un petit chromosome comme le chr 21 n'en représente que 1%. Ainsi, une mégabase du chromosome 21 a donc plus de chance d'interagir avec les autres chromosomes (qui représentent 99% du volume nucléaire) qu'une mégabase du chromosome 1 n'a de chance d'interagir avec les autres chromosomes (qui représentent 92.5% du volume nucléaire). Ceci expliquerait l'augmentation du *trans*-index lorsque la taille du chromosome diminue.

Ce graphique nous apprend aussi que la proportion d'interactions en *trans* est comprise entre 54 et 62% (extrémité des barres d'erreurs) pour le chromosome 5. Ces valeurs sont significativement plus élevées que le pourcentage d'interactions en *trans* observé par 4C pour les 10 points de vue du chr 5 dans le même type de cellules (20 à 30% dans des lymphoblastoïdes, tableau 2). Il est envisageable que, dans chaque boîte d'une Mb, il y ait à la fois des loci établissant beaucoup de *trans*-contacts et d'autres beaucoup moins, et que les points de vue choisis pour l'analyse 4C fasse justement partie de ceux établissant moins de *trans*-interactions.

L'analyse des données Hi-C montre que la proportion d'interactions en *trans* de chaque chromosome dépend d'abord de sa taille : plus le chromosome est petit, plus, pour chaque Mb, la proportion d'interactions impliquant un autre chromosome est élevée.

b- Des variations du *trans*-index ($\pm 10\%$) sont néanmoins visibles

Deux régimes de variations du *trans*-index

Si le *trans*-index est en moyenne homogène le long de chaque chromosome, de légères variations de celui-ci sont néanmoins visibles. On observe deux régimes :

- un régime basse-fréquence (variations sur plusieurs dizaines de Mb) : le *trans*-index a tendance à diminuer à mesure que l'on s'éloigne du centromère du chr 1 (notamment sur le bras p) et du chr 21 (excepté une remontée proche du télomère) (figure 53).
- un régime haute-fréquence (variations sur quelques Mb) : la région autour de 50 Mb sur le chr 1 établit plus d'interaction en *trans* que ses voisins immédiats ; à l'inverse la région autour de 90 Mb sur le chr 12 établit moins d'interaction en *trans* que ses voisins immédiats. Ces oscillations visibles dans le *trans*-index ont une période de l'ordre de 5-10 Mb.

L'hétérochromatine tend à interagir un peu plus en *trans*

Afin de rechercher une explication génétique, épigénétique, et/ou fonctionnelle à ces variations du *trans*-index, nous avons tenté de corréliser ces deux régimes d'oscillations avec des marques génomiques (teneur en GC), épigénétiques (H4K4me3, H3K36me3, H3K27me3 et H3K9me3) et avec le timing de réplication (figure 54).

- A mesure que l'on s'éloigne du centromère du chr 1 sur le bras p (régime à basse-fréquence), la diminution du *trans*-index semble être associée à une augmentation de la teneur en GC, une augmentation de la proportion d'euchromatine, et un timing de plus en plus précoce. Le bras q du chr 12 semble suivre la même tendance (au moins pour la teneur en GC et la proportion d'euchromatine).
- Certains maxima locaux du *trans*-index (régime haute-fréquence) coïncident avec des zones localement enrichies en H3K9me3, pauvres en GC et à timing tardif (par exemple sur le chr 1, celui à 50 Mb ou 105 Mb).

Ces deux observations indiquent que les régions de chromatine inactive établissent plus de *trans*-contacts que les régions de chromatine active.

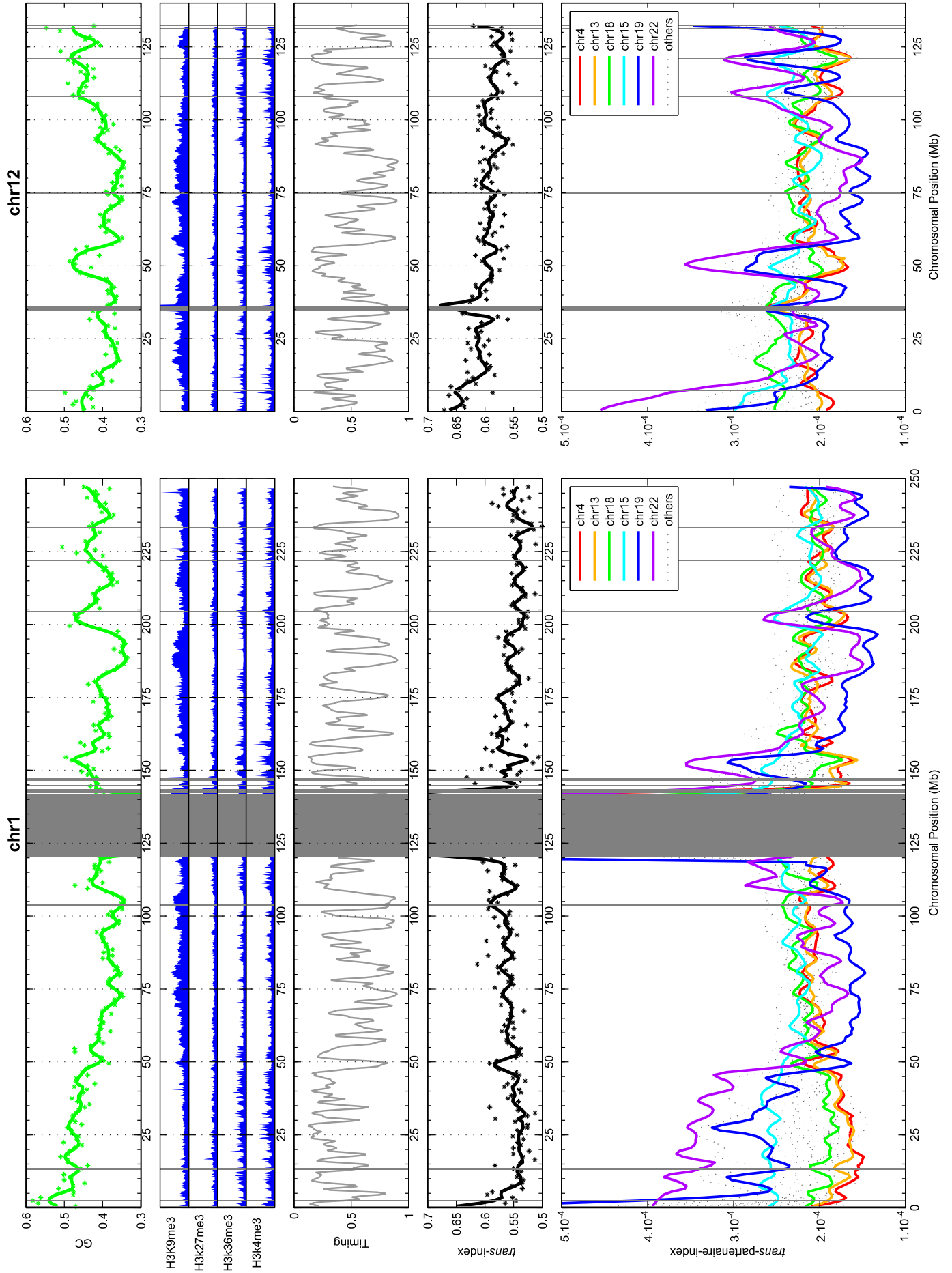
Néanmoins, le lien entre *trans*-interactions fréquentes et hétérochromatine n'est pas direct

La corrélation entre la capacité à établir des interactions en *trans* et le type de chromatine n'est pas simple :

- Sur le bras p du chr 12, alors que la teneur en GC et la proportion d'euchromatine augmentent (régime à basse-fréquence), et que le timing devient de plus en plus précoce en s'éloignant du centromère, le *trans*-index augmente également. Sur ces 30 Mb du chr 12, il semblerait donc que ce soit l'euchromatine qui établisse plus de *trans*-contacts, alors qu'il s'agissait de l'hétérochromatine sur les 50 premiers Mb du chr 1.
- De plus, si la correspondance entre euchromatine et *trans*-contact était directe, deux régions possédant le même *trans*-index devraient être situées dans le même contexte chromatinien. Or les loci autour de 70 Mb et de 115 Mb sur le chr 12, qui ont le même *trans*-index, ne possèdent ni la même teneur en H3K9me3 ou H3K4me3, ni la même teneur en GC, ni le même timing de réplication.
- Enfin, la zone d'euchromatine autour de 50 Mb du chr 12 (maximum en GC, déplétion en H3K9me3, présence de H3K4me3 et timing précoce) n'est pas associée à une variation claire du *trans*-index qui oscille autour des mêmes valeurs qu'au niveau des voisins hétérochromatiques (figure 54).

Des variations du *trans*-index existent, indiquant que certaines portions de chromosomes sont capables d'établir plus de *trans*-contacts que leurs voisins immédiats. Si, sur le chr 1, il semble que l'hétérochromatine est capable d'établir plus de *trans*-contacts que l'euchromatine, cette tendance n'est pas partagée par tous les chromosomes.

FIGURE 54 – *Trans*-partenaires le long des chromosomes 1 et 12. La proportion d'interactions en *trans* le long du chromosome 1 et 12 (panneau central, noir) a été comparée avec des marques génomiques (teneur en GC, vert) échantillonnées à 1 Mb et épigénétiques (H3K9me3 -PBMC- et H3K27me3/K36me3/K4me3 -Gm12878-, bleu) échantillonnées à 200 kb. Le timing de réplication (moyenné sur 1 Mb, avec un pas glissant de 400 kb) est représenté par la courbe grise. Afin de détecter la présence éventuelle de chromosomes partenaires préférentiels, le *trans*-partenaire-index (voir méthode et texte) a été évalué le long des chr 1 et 12. Les chr 4, 13, 18, 15, 19 et 22 sont représentés en traits pleins colorés, tandis que les autres sont représentés en pointillés gris (seul le lissage par régression polynomiale locale excluant les points situés à plus de 6 écarts-types est représenté). Les chr 4, 13 et 18 sont les trois autosomes les plus tardifs (timing de réplication de l'ensemble du chromosome) et ont une teneur en GC basse ; les chr 15, 19 et 22 font partie des 4 chromosomes les plus précoces et ont une teneur en GC élevée.



Notre observation diffère des résultats de Kalhor et al. [2012]

Ce résultat n'est pas en accord avec l'étude de Kalhor et al. [2012] dans laquelle il est démontré, grâce à une expérience de TCC (Tethered Conformation Capture), que les régions de chromatine active établissent plus de contacts inter-chromosomiques que les régions de chromatine inactive. Le TCC, proche du Hi-C, cherche à optimiser la détection de *trans*-contacts grâce à une ligation s'effectuant sur support solide (les fragments de restrictions cross-linkés sont ancrés sur des billes avant la ligation). Selon les auteurs du TCC, cet ancrage sur support solide limite la formation de ligations aléatoires entre deux fragments de restriction non cross-linkés. Or, ces ligations aléatoires ont plus de chance de survenir entre deux chromosomes différents, qu'au sein d'un même chromosome. Les ligations aléatoires ont donc tendance à augmenter la fréquence des *trans*-contacts. Avec l'ancrage sur billes, Kalhor et al. [2012] ont montré que la proportion de ligations en *trans* est diminué par 2 : la librairie TCC ne contient que 25-30% de *trans*-ligations alors que la librairie Hi-C en contient plus de 50%. Ce pourcentage est d'ailleurs nettement plus compatible avec celui observé en 4C (20 à 30% pour le même type de cellules –lymphoblastoïdes–, table 2). Il est donc possible que les *trans*-ligations soient surreprésentées dans le Hi-C, compliquant leur analyse.

5.4.3- Chromosomes préférentiellement contactés

S'il n'y a pas de variations de grande ampleur du *trans*-index, ceci ne signifie pas que les chromosomes contactés sont toujours les mêmes. Pour visualiser d'éventuels chromosome-partenaires privilégiés, nous avons défini un *trans*-partenaire-index qui quantifie la proportion de *trans*-interactions impliquant un chromosome donné. Puisqu'un grand chromosome a plus de chance d'interagir avec lui-même (figure 53), les *trans*-interactions seront aussi biaisées en faveur des grands chromosomes (le chr 10 aura plus de chance de contacter le chr 1 qui représente 7,5% du volume nucléaire que le chr 21 qui n'en représente qu'1%). Pour pouvoir comparer la capacité à interagir plus préférentiellement avec un chromosome plutôt qu'un autre, l'index défini normalise les *trans*-interactions par la taille du chromosome contacté (voir méthode).

Les profils sont représentés figure 54 (panneau du bas). Seuls 6 chromosomes sont représentés en trait pleins. Les chr 4, 13 et 18 sont les trois autosomes les plus tardifs (timing moyen de la totalité du chromosome) et ont une teneur en GC comprise entre 38.2% et 40.1%. A l'inverse, les chr 15, 19 et 22 font partie des 4 chromosomes les plus précoces et ont un GC moyen compris entre 45.5% et 48.4%.

Avec cette analyse, on observe clairement que les portions de chromosomes ayant un GC élevé et un timing précoce sur plusieurs Mb (sur le chr 1 : les 50 premiers Mb, à 150 et 200 Mb ; sur le chr 12 : les 10 premiers Mb, à 50, 110 et 120 Mb) contactent préférentiellement les chromosomes ayant, eux aussi, un timing moyen précoce et une teneur en GC élevée. A l'inverse, les régions localement plus pauvres en GC et à timing tardif (chr 1 : autour de 190 Mb et chr 12 : autour de 85 Mb) contactent moins les chromosomes à haut GC et à timing précoce (chr 15, 19 et 22).

Curieusement, les régions localement très précoces et montrant cette forte interaction préférentielle avec les autres chromosomes précoces peuvent avoir un *trans*-index localement plus faible (chr 1 : les 50 premiers Mb et à 150 Mb ;

chr 12 à 50 Mb). Il s'agirait donc de régions interagissant globalement moins avec d'autres chromosomes, mais lorsqu'elles le font, elles contactent préférentiellement un chromosome à timing précoce et riche en GC.

Dans le noyau, ces grandes régions précoces, comme les 50 premiers Mb du chr 1, sont donc proches des chromosomes répliqués en début de phase S. Ce positionnement moyen n'exclut pas que les quelques loci tardifs retrouvés dans ces grandes régions puissent préférentiellement interagir avec d'autres loci tardifs (y compris en *trans*, comme le laisse envisager le 4C sur le chr 5). Néanmoins, à l'échelle d'analyse proposée, l'influence de ces rares loci tardifs isolés sur les *trans*-partenaires moyens de la région n'est pas visible.

Conclusion des résultats de l'analyse Hi-C

Notre analyse des *trans*-interactions dans le jeu de données Hi-C montre que certains loci sont capables d'établir un peu plus de *trans*-contacts que leurs voisins. Néanmoins, nous n'avons pas pu préciser si les bordures du territoire chromosomique correspondent plutôt à de l'euchromatine (bras p du chr 12) ou à de l'hétérochromatine (bras p du chr 1). Il est possible que les régions en surface du territoire chromosomique ne fassent que 100 kb de long et soient dispersées le long du chromosome. Ainsi, leur influence sur la valeur du *trans*-index, échantillonné à 1 Mb, pourrait être gommée. Dans tous les cas, le résultat de notre analyse est différent de celui obtenu avec les données TCC à partir desquelles les auteurs ont proposé que les régions de chromatine ouverte établissent plus de *trans*-contacts et se localisent en périphérie (voire hors) du territoire chromosomique.

Notre analyse a néanmoins permis de montrer que les régions qui sont riches en GC et précoces sur plusieurs Mb (au moins 5 Mb) *trans*-contactent préférentiellement des chromosomes qui sont eux-aussi précoces et riches en GC (chr 15, 19, 22). A l'inverse, les régions pauvres en GC et tardives sur plusieurs Mb *trans*-contactent peu ce jeu de chromosomes.

Conclusion du chapitre

Au cours de ce chapitre, différents liens entre la topologie de la chromatine et le timing de réplication ont été caractérisés.

- ✓ Premièrement, comme pour les cellules lymphoblastoïdes, nous avons démontré que la chromatine des cellules K562 (cellules pro-érythrocytaires) est organisée de telle sorte à ce que les interactions entre loci répliqués au même moment de la phase S sont favorisées. De manière frappante, la comparaison des 4C effectués dans les cellules K562 et lymphoblastoïdes montre que les contacts à longue-distance sont identiques dans les deux types cellulaires lorsque le timing est conservé mais divergent dès lors que le timing de réplication est différent.
- ✓ Deuxièmement, notre analyse 4C révèle qu'au sein d'un plateau précoce, des interactions préférentielles s'établissent entre loci partageant le même contexte chromatinien (mêmes modifications post-traductionnelles d'histones, mêmes facteurs de transcription, présence de CTCF et de cohésines).

- ✓ Troisièmement, nous avons montré que ce plateau précoce établit des contacts à longue-distance similaires à ceux que pourraient établir un pic de timing. La ségrégation des régions précoces et tardives dans l'espace nucléaire est donc générale et ne se limite pas aux pics de timing.
- ✓ Enfin, l'analyse des données Hi-C révèle qu'au sein d'une région en moyenne précoce sur plusieurs Mb (au moins 5-10), les interactions inter-chromosomiques favorisent les chromosomes répliqués en début de phase S et possédant une haute teneur en GC.

Partie C

Discussion

C

Discussion et conclusion

Comprendre et identifier les déterminants structuraux et fonctionnels qui contrôlent ou modulent l'agencement de la chromatine dans le noyau est un but essentiel. Si ce travail n'a pas pour vocation d'apporter une réponse définitive à ces questions fondamentales, il s'inscrit dans une stratégie d'identification de domaines topologiques dans le noyau. Mon travail a permis de souligner l'existence de domaines structuraux dont les frontières semblent avoir un lien étroit avec le timing de réplication.

6.1- L'organisation de la chromatine à grande échelle est corrélée au timing de réplication à différentes échelles

Grâce à une approche 4C couplée à un séquençage haut-débit, les partenaires chromatinien de loci répliqués en début et en fin de phase S ont été identifiés. Cette étude nous a permis de souligner l'existence d'un lien entre la topologie de la chromatine et l'ordre dans lequel différents loci sont répliqués pendant la phase S. Ce lien a été observé à des échelles allant de la mégabase au chromosome (voire génome) entier.

6.1.1- Conformation à l'échelle de quelques mégabases

Interaction et proximité des régions précoces Deux loci précoces consécutifs se contactent l'un l'autre. Ce contact préférentiel a été observé chez des cellules lymphoblastoïdes, des PBMC (cellules mononucléées du sang périphérique) en cycle ou dormantes (chapitre 4), ainsi que dans des cellules K562 (issues d'une leucémie myéloïde, chapitre 5). Cette interaction correspond à un premier niveau de repliement ayant un lien avec la réplication. Les mesures de distance par FISH démontrent qu'un locus précoce est, dans l'ensemble, plus proche d'un autre locus précoce que d'un locus tardif (chapitre 4). Il est donc probable que cette interaction repose sur la formation de boucles géantes de chromatine contenant

les régions tardives intermédiaires. Cette interaction entre voisins immédiats est confirmée par l'analyse du jeu de données Hi-C proposée figure 46.

De façon remarquable, cette interaction entre régions précoces ne se limite pas au plus proche voisin. Sur une dizaine de Mb, un point de vue précoce interagit avec l'ensemble des séquences précoces alentours. Ces interactions sont ponctuelles puisque le rebond visible dans le profil 4C a une largeur de l'ordre de la centaine de kilobase. Un plateau précoce (séquence présentant, sur quelques Mb, un timing de réplication homogène avec une valeur proche de 0) est également capable d'interagir avec les quelques voisins précoces (même si ses interactions semblent moins contraintes, chapitre 5).

Le maintien des interactions entre séquences précoces pourrait reposer sur la protéine CTCF. En effet, ce facteur impliqué dans l'agencement tridimensionnel de la chromatine, est plutôt retrouvé dans les régions précoces [Ryba et al., 2010], et particulièrement au niveau des pics de timing [Baker et al., 2012], notamment ceux du chromosome 5 qui ont été la cible d'amorces 4C.

Une vallée de timing forme un domaine structural dans le noyau Les régions tardives ont un comportement sensiblement différent. Un point de vue situé au cœur d'une vallée de timing est capable d'interagir avec l'ensemble des séquences de la vallée, avec à peu près la même fréquence (voir notamment les profils 4C obtenus avec V4, figures 37 - 39 - 48). Cette capacité d'interaction apparaît brusquement abolie dès qu'un pic de timing doit être franchi. Au niveau des pics de timing, la chromatine est donc organisée de sorte à limiter la propagation d'interactions fortes impliquant une vallée de timing. Cette frontière physique, quelle que soit la manière dont elle est spécifiée, est très clairement visible sur les matrices Hi-C échantillonnées à 100 kb (figure 47). Au vu de ces résultats, on en conclut que les U-domaines (une vallée bordée par deux pics de timing, page 75) forment des domaines structuraux au sein desquels de multiples interactions peuvent s'établir. Les domaines fonctionnels identifiés à partir du timing de réplication possèdent donc des homologues structuraux. CTCF qui, on l'a dit, est présent au niveau des pics de timing, est donc un candidat potentiel pour spécifier ces frontières.

Grâce à un Hi-C à haute-résolution effectué dans des cellules souches humaines et murines, l'étude de Dixon et al. [2012] souligne l'existence de "domaines topologiques" couvrant 90% du génome. D'une taille légèrement inférieure à 1 Mb, ces domaines sont relativement isolés les uns des autres, et leurs frontières corrélerent avec la présence de CTCF et/ou la présence de gènes de ménage. Pour déterminer si ces "domaines topologiques" [Dixon et al., 2012] et les "U-domaines structuraux" correspondent aux mêmes objets, j'ai comparé, pour la portion [150 - 170 Mb] du chromosome 5, le timing de réplication de cellules souches humaines (cellules H1, <http://www.replicationdomain.com/>) avec les matrices d'interactions du Hi-C effectuées dans les mêmes cellules (<http://chromosome.sdsc.edu/mouse/hi-c/>). Cette comparaison (non exposée dans ce manuscrit) a montré que la majorité des pics de timing coïncident avec les frontières de "domaines topologiques" identifiés par Dixon et al. [2012]. Plus exactement, les plus grands U-domaines de timing seraient composés de plusieurs "domaines topologiques". Les objets identifiés dans cette étude correspondraient alors à une subdivision des U-domaines. Cette imbrication en poupées russes est d'ailleurs visible dans les données Hi-C de cette étude, puisqu'en modifiant la palette de couleurs (<http://chromosome.sdsc.edu/mouse/hi-c/>), on s'aperçoit clairement que plusieurs "domaines topologiques" s'agglomèrent en de plus grands domaines dont les frontières coïncident avec les pics de timing. Cette

exploration, qui demande évidemment à être confirmée et quantifiée sur l'ensemble du génome, souligne vraisemblablement l'existence de niveaux de repliements hiérarchiques dans la chromatine : un domaine d'auto-interaction peut vraisemblablement être inclus dans un autre.

Interaction des régions tardives En plus de former des domaines structuraux bordés par les pics de timing, les vallées de timing V3 et V4 sont capables de contacter d'autres vallées de timing, aussi bien dans les cellules lymphoblastoïdes que dans les PBMC (figures 37 - 39). Les interactions établies par V3 et V4 se répartissent sur une région comprenant la quasi-totalité des vallées de timing voisines (accolades, figures 37 et 39). Les contacts à quelques Mb de distance semblent donc moins contraints pour les vallées que pour les pics de timing qui possèdent des partenaires ponctuels (cf. avant). En étant moins contraintes, les interactions ne s'établissent donc pas forcément entre les deux loci ciblés par des sondes FISH. Ceci pourrait expliquer pourquoi, malgré les 300 mesures de distances effectuées, les contacts préférentiels entre loci tardifs ne s'accompagnent pas d'une proximité physique qui pourrait être vue par FISH (figure 45).

Il est possible que les interactions entre séquences tardives s'établissent par défaut. En effet, la proximité physique marquée concerne les pics de timing au niveau desquels se trouve CTCF. Il est possible que ce facteur verrouille la proximité spatiale entre les pics de timing, les isolant en même temps des régions tardives intermédiaires (ceci expliquerait la chute brutale de la fréquence d'interaction des points de vue de type "V" lorsqu'un pic de timing doit être franchi). En rendant les régions précoces "inaccessibles" aux régions tardives, quelle que soit la manière dont est spécifiée cette isolation, les seuls partenaires possibles des régions tardives sont les autres régions tardives. Ce contact par défaut serait statistiquement détecté par 4C (reflet de dizaines de milliers de cellules), mais pas par FISH (tout au plus, deux régions tardives ne sont jamais très loin l'une de l'autre, vision sur une centaine de cellules). Ce contact par défaut s'accorde d'ailleurs bien avec l'absence de partenaires chromatiniens ponctuels pour les régions tardives.

Une représentation possible d'une chaîne de globules tardifs Une représentation possible de ces contacts à quelques mégabases est avancée figure 55A. A la fin de la première section du chapitre 1, nous avons représenté l'enroulement de la fibre de 30 nm sous forme de boucles. Notre analyse 4C permet de préciser comment celles-ci sont arrangées. Je propose que cette succession de boucles soit ponctuellement interrompue par la présence de pics de timing (bleu) au niveau desquels on retrouve CTCF. Ces régions précoces interagissent ensemble et sont physiquement proches les unes des autres. Les régions tardives intermédiaires (jaune) forment un "globule" dont l'agencement n'est pas décrit dans cette étude. Il pourrait tout à fait s'agir d'un globule fractal puisque la gamme d'échelle sur laquelle J. Dekker et collaborateurs proposent qu'il existe (500 kb → 5 Mb [Lieberman-Aiden et al., 2009; Zhang et al., 2012]) est compatible avec la taille des vallées de timing (la vallée ciblée par V4 fait 4 Mb de long). Mais il faudrait également tenir compte de l'imbrication de sous-domaines d'auto-interactions (voir avant et Dixon et al. [2012]). Mon travail ne cherchant pas à préciser la structure du "globule" formé par les vallées de timing, celui est représenté figure 55 sous forme de spirales. Cette organisation globale s'accorde avec la formation de contacts au sein d'une même vallée de timing (flèche A1), avec la formation de contacts entre régions précoces qui sont proches les unes des autres et où se trouve CTCF

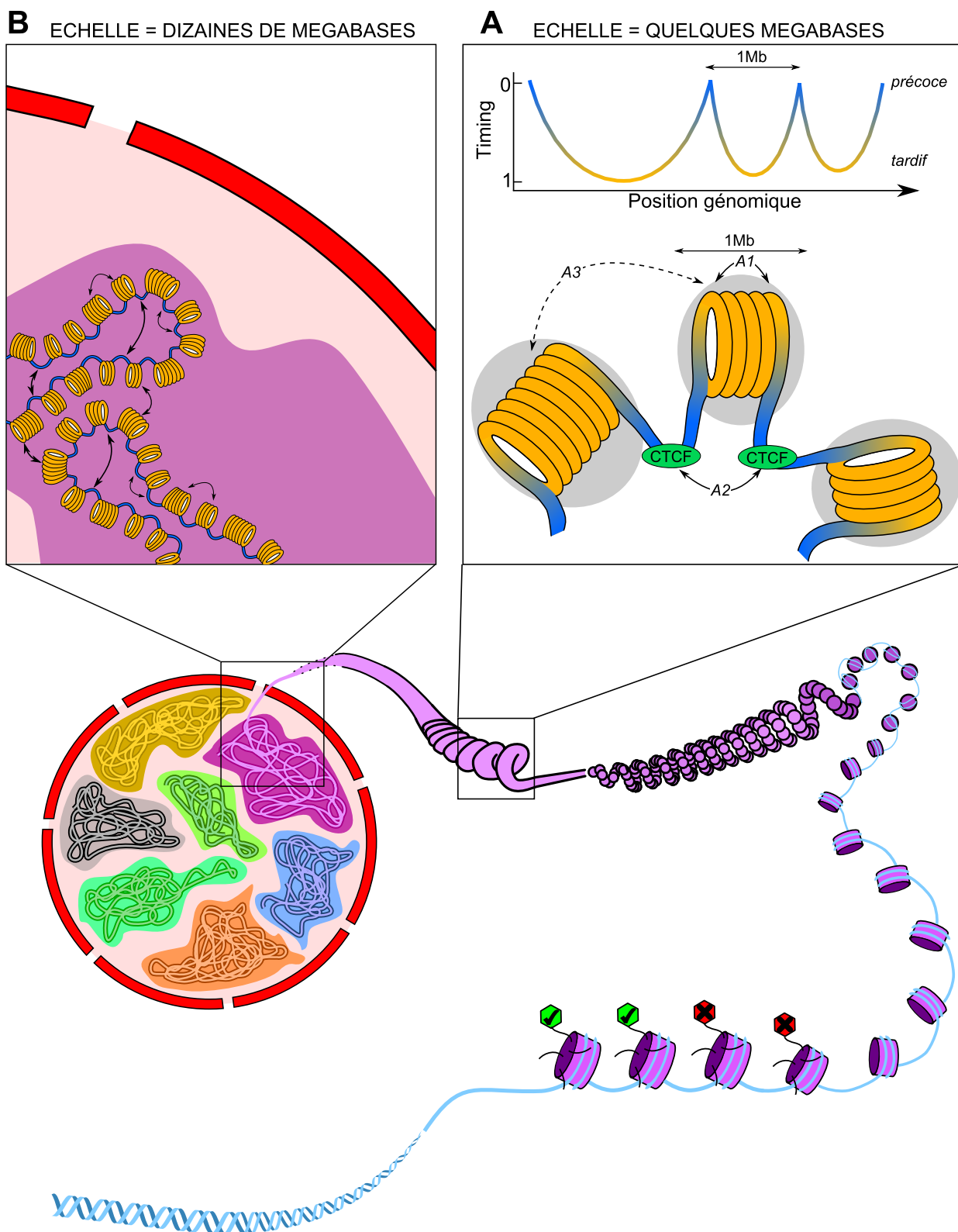


FIGURE 55 – Organisation topologique de la chromatine en suivant le timing de réplification. Ce schéma reprend celui de la figure 9. Nous pouvons maintenant détailler l'agencement des boucles géantes de chromatine. **A-** Les régions répliquées en fin de phase S (jaune) forment des globules d'auto-interaction (zone grisé, flèche A1), représentés ici en spirale, et les régions précoces (bleu) sont retrouvées à la jonction de ces globules. Les régions précoces interagissent ensemble et sont physiquement proches dans l'espace (flèche A2). Elles sont isolées des séquences tardives par CTCF; les seules partenaires possibles des séquences tardives sont alors les séquences tardives voisines (flèche A3). **B-** Le repliement de la chromatine à plus grande échelle favorise les interactions entre loci précoces d'un côté, et loci tardifs de l'autre.

(flèche A2), ainsi qu'avec la formation de contacts moins contraints entre régions tardives consécutives (flèche A3).

6.1.2- Conformation à l'échelle de plusieurs dizaines de Mb

Cette interaction préférentielle entre régions précoces d'un côté et régions tardives de l'autre est retrouvée pour des séparations génomiques beaucoup plus grandes, allant de quelques dizaines de Mb à la centaine de Mb (taille du chromosome entier). Cette ségrégation de la chromatine en fonction du timing de réplication est observée dans des cellules lymphoblastoïdes et PBMC (chapitre 4), ainsi que dans des cellules K562 (chapitre 5).

A cette échelle, nos travaux sont en accord avec des résultats précédents montrant, par l'étude des données Hi-C, que des séquences répliquées de manière synchrone interagissent plus fréquemment qu'attendu [Yaffe et al., 2010; Ryba et al., 2010; De and Michor, 2011]. Cependant, grâce à un nombre élevé de lectures pour chaque point de vue, notre 4C autorise la visualisation directe des interactions entre séquences précoces et des interactions entre séquences tardives, ce qui n'était pas possible avec les données Hi-C (voir pages 88 et suivantes). Il permet ainsi d'affiner cette conclusion. Nous avons en effet montré que chaque pic de timing est capable d'établir une interaction avec toutes les autres séquences précoces du chromosome entier. Ces interactions s'établissent avec une extraordinaire spécificité puisque les profils 4C des points de vue de type "P" suivent admirablement le timing de réplication (chapitres 4 et 5).

Notons qu'à cette échelle, un plateau précoce s'associe préférentiellement avec d'autres séquences précoces. Il ségrège donc, dans l'espace nucléaire, avec les pics de timing (chapitre 5).

Une représentation possible de ces contacts à quelques dizaines de Mb est avancée figure 55B. Dans celle-ci, le repliement permet les contacts entre régions précoces (bleu) et entre régions tardives (jaune), sans cross-contacts possibles.

6.1.3- Conformation à l'échelle du territoire chromosomique

Les interactions inter-chromosomiques respectent les "règles" d'interaction vues au sein d'un chromosome : les points de vue précoces P1, P2, P3 et P4 *trans*-contactent plutôt des séquences précoces, et les points de vue tardifs (V1, V3 et V4) plutôt des séquences tardives (figure 44).

L'analyse des données Hi-C effectuée au chapitre 5 apporte un éclairage supplémentaire sur la manière dont s'organisent les contacts inter-chromosomiques. Les portions de chromosomes, qui sont en moyenne précoces sur plusieurs mégabases, établissent très majoritairement des *trans*-interactions avec les autres chromosomes répliqués en début de phase S (figure 54). Ceux-ci correspondent majoritairement à des petits chromosomes riches en GC (chr 19, chr 22 par exemple) que l'on retrouve préférentiellement au centre du noyau (voir page 35 dans l'introduction). Ceci signifie que les grandes régions précoces sur plusieurs Mb (comme les 50 premiers Mb du chr 1, ou sur le chr 12 autour de 50 Mb) adoptent, dans le territoire chromosomique, un positionnement radial qui les rapproche du centre du noyau. A l'inverse, les portions de chromosomes qui sont en moyenne tardives sur plusieurs Mb, *trans*-contactent peu ces petits chromo-

somes riches en GC et précoces : dans le territoire chromosomique, elles adoptent donc vraisemblablement un positionnement radial les éloignant du centre du noyau.

6.2- Les profils d'interactions sont mieux décrits par le timing que par le niveau de transcription

Un des résultats marquant de notre étude 4C est l'association dans l'espace nucléaire des régions précoces d'un côté, et des régions tardives de l'autre. Cette ségrégation dans l'espace nucléaire est en accord avec des résultats précédents obtenus par d'autres groupes qui montrent, par des approches de type 4C ou Hi-C, l'absence relative de contacts entre la chromatine active (précoce) et la chromatine inactive (tardive) [Simonis et al., 2006; Lieberman-Aiden et al., 2009; Sexton et al., 2012; Zhang et al., 2012].

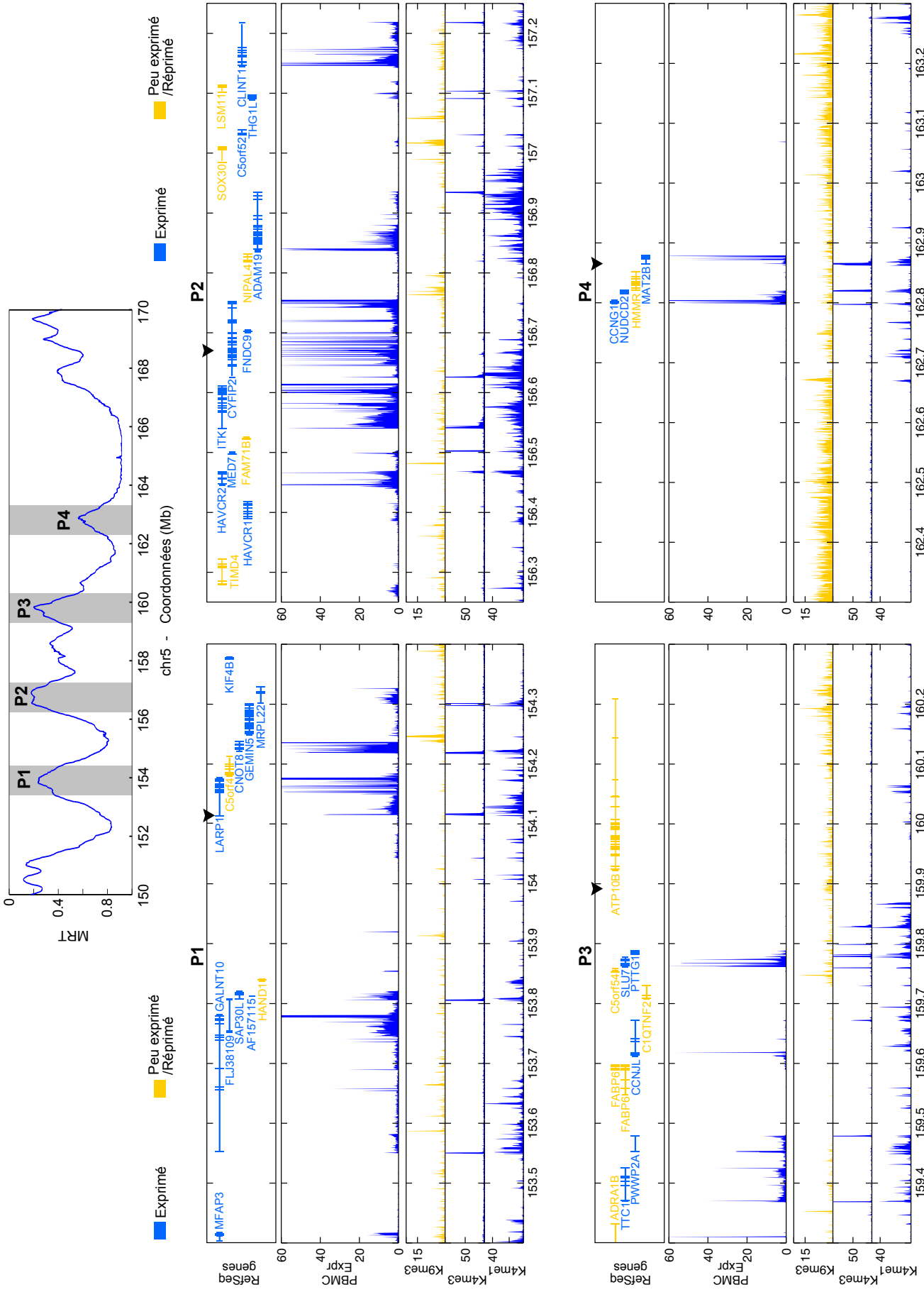
Une question clé qui demeure est de déterminer quel processus nucléaire contraint (ou résulte de) cette ségrégation. En effet, les régions précoces sont majoritairement riches en gènes transcrits, en acétylation d'histones, en H3K4me3 et H3K36me3, alors que les régions tardives sont, à l'inverse, pauvres en gènes, hypo-acétylées, riches en domaines d'association à la lamina nucléaire et riches en H3K9me3 (figure 20 du chapitre 2). Par conséquent, il peut sembler difficile d'avancer l'existence d'un lien entre la fréquence de contacts et le timing de réplication, plutôt qu'avec l'expression génique. Je propose, dans ce contexte, trois pistes. Chacune tente d'aborder cette question par une approche différente.

Le taux de transcription ne suffit pas pour intégralement prédire les contacts à longue-distance puisque les partenaires des points de vue de type "P" comprennent à la fois des gènes exprimés et des gènes réprimés.

Si le taux de transcription est l'unique paramètre conduisant à une ségrégation nucléaire de la chromatine active et de la chromatine inactive, alors les points de vue P1, P2 ou P4 (situés au sein d'un gène exprimé) devraient être physiquement plus proches de gènes exprimés que de gènes réprimés.

Or, les partenaires les plus fréquents des points de vue "P", à savoir les quelques pics de timing alentour, possèdent à la fois des gènes exprimés (bleu) et des gènes réprimés (jaune, voir figure 56). Ceci signifie que lorsque P1 contacte P4, il se retrouve à la fois proche physiquement de gènes fortement exprimés (CCNG1, NUDC2, MAT2B), et de gènes réprimés (HMMR, figure 56). Un même constat peut-être dressé pour les autres pics de timing. Ceci signifie, qu'à la résolution du 4C proposée dans ce manuscrit, on ne peut pas conclure que le point de

FIGURE 56 – Le taux de transcription n'explique pas à lui-seul les contacts chromatiniens. Présence de gènes (RefSeq), taux d'expression et teneur en H3K9me3, H3K4me3 et H3K4me1 (voir figure 36 pour la provenance des données) sur un intervalle de 1 Mb centré sur les pics de timing ciblés par P1, P2, P3 et P4 (P5 suit la même tendance). Le niveau d'expression (RNA-seq, GEO GSM669616), et les profils d'histone ont été obtenus dans des PBMC dormantes humaines. Les gènes (panneau RefSeq) sont colorés en fonction de leur niveau d'expression. Le timing (haut centre) correspond à celui des cellules gm06990 [Hansen et al., 2010; Baker et al., 2012]. Les données d'expression et épigénétiques ont été établies sur l'assemblage hg19 et rétro-converties en hg18.



vue P1 est plus proche d'un gène exprimé que d'un gène réprimé. Le niveau de transcription d'un gène isolé ne permet pas à lui seul de prédire l'établissement d'associations préférentielles.

Par ce raisonnement, on retrouve une propriété connue du timing de réplication qui a déjà été abordée en introduction (page 72) : la corrélation entre le timing de réplication et le taux de transcription n'est forte que lorsqu'elle est intégrée sur de grandes fenêtres, de l'ordre de 600 kb chez la souris [Hiratani et al., 2008] (et probablement aussi chez l'homme). Un timing de réplication précoce reflète la contribution de plusieurs gènes actifs, autour desquels on trouve aussi des gènes réprimés. Le timing de réplication "récapitule", en quelque sorte, cet état moyen de la chromatine. Avec la résolution du 4C proposée dans ce manuscrit, c'est cet "état moyen" de la chromatine, résumé par le timing de réplication, qui paraît lié à l'établissement de contacts préférentiels. Une analyse plus fine (à l'échelle de la dizaine de kb) permettrait, peut-être, d'identifier précisément les partenaires des points de vue de type "P" qui ancrent les interactions entre pics de timing et de confirmer / infirmer le rôle éventuel de la transcription.

Des regroupements de séquences co-répliquées (foyers de réplication) existent en absence de transcription

L'observation de foyers de réplication dans le micro-noyau de *Stylonychia lemnae* est un argument supplémentaire pour l'existence d'une organisation de la chromatine liée à la réplication plutôt qu'à la transcription. En effet, le micro-noyau de *Stylonychia* est transcriptionnellement inerte, avec des marques répressives d'histones (H3K27me3) mais pas de marques actives (pas d'acétylation, ou de H3K4me1/3) [Postberg et al., 2008]. Celui-ci (avec les 2×10^9 bp qu'il renferme) contient des foyers de réplication. La taille de ces foyers est similaire à celle mesurée chez les métazoaires (0.5 μ m) et surtout, ils s'activent en suivant un programme spatio-temporel qui rappelle celui observé chez les mammifères (de petits spots au centre du noyau en début de phase S, suivi de spots plus gros en périphérie à la fin de la phase S [Postberg et al., 2005]. On assiste donc à une ségrégation physique des loci précoces et tardifs. Si la stricte correspondance entre les foyers de réplication et les contacts chromatiniens capturés par 4C reste à établir, il est probable que certains foyers contiennent des fourches issues de plusieurs pics de timing consécutifs. En effet, puisque chaque foyer contient des séquences se répliquant de manière synchrone, il doit donc s'agir, en tout début de phase S, de séquences situées sur des pics de timing voisins. La présence de tels foyers dans le micro-noyau de *Stylonychia*, alors que celui-ci est inerte du point de vue de la transcription, suggère fortement que des contacts chromatiniens entre régions répliquées de manière synchrone peuvent s'établir en absence de toute transcription.

Les influences de la transcription et de la réplication pourraient se cumuler pour favoriser les contacts

Un haut niveau de transcription pourrait potentialiser les interactions à longue-distance. En effet, les partenaires du point de vue P3 sont moins corrélés au timing de réplication que ceux de P1, P2 et P4 (figure 43 du chapitre 4). Or, à la différence de ces trois derniers, P3 est plus proche d'un gène non-exprimé (ATP10B, figure 56) que d'un gène exprimé (le gène fortement exprimé le plus proche – SLU7 – est à 100 kb du point de vue). Puisque que P3 interagit avec d'autres pics de timing, mais de manière moins stricte que P1, P2 et P4, ceci suggère que les influences de la transcription et de la réplication pourraient se cumuler pour favoriser les contacts à longue-distance, éventuellement par un an-

cage au niveau des foyers de réplication et de transcription qui, selon l'étude de [Malyavantham et al. \[2008\]](#), sont agencés l'un à proximité de l'autre.

Seuls quelques contacts à quelques Mb de distance ne semblent pas s'établir entre régions répliquées simultanément ou présentant le même niveau d'expression. En effet, dans les cellules lymphoblastoïdes (figure 37 et figure 48) et les PBMC (dormantes ou activées, figure 39), P4 interagit avec une région située juste en aval de 164 Mb. Pourtant, ce partenaire ne correspond ni à un pic de timing, ni à une région exprimée, ni à une unité de transcription connue et aucune marque épigénétique de chromatine active n'est clairement présente. CTCF et les cohésine sont également absents. Ce contact préférentiel de P4 avec un partenaire situé 1 Mb en aval n'est donc ni soutenu par la réplication, ni par la transcription, ni par la présence de marques épigénétiques. Indéniablement, la transcription et la réplication ne permettent pas de contraindre l'intégralité des contacts chromatinien.

Dissocier les influences des différents processus nucléaires (réplication, transcription, mais aussi réparation) pour élucider le(s) déterminant(s) expliquant au mieux l'agencement de la chromatine est assurément l'un des défis à venir. Ces différentes fonctions étant imbriquées (sauf sur le micro-noyau des protozoaires ciliés), l'étude de l'agencement de la réplication et de la transcription dans le volume nucléaire pourrait apporter un nouvel éclairage sur cette question. L'utilisation de techniques de microscopie optique super-résolutive avec un marquage à la fois de la réplication (dérivés de desoxyuridine –BrdU, CldU, IdU, EdU–, ou protéines fusions de type PCNA-GFP) et de la transcription (dérivés d'uridine, immunofluorescence dirigée contre l'ARN polymérase phosphorylée sur la sérine 5) pourrait apporter un éclairage supplémentaire sur cette question.

En conclusion, à l'échelle d'analyse proposé dans ce manuscrit, les contacts chromatinien sont mieux prédits par le timing de réplication que par le niveau de transcription d'un gène isolé. Quelle que soit la séparation génomique, une séquence précoce aura pour partenaire préférentiel une autre région précoce plutôt qu'une région tardive situées à la même distance. Ce résultat est valable pour tous les types cellulaires utilisés dans ce manuscrit (cellules lymphoblastoïdes, PBMC dormantes ou en cycle, ou cellules K562).

6.3- Des contacts chromatinien liés au timing de réplication sont observés dans des cellules en phase G0

Les liens entre organisation de la chromatine à grande échelle et la réplication sont également observés dans des cellules en phase G0 (PBMC dormantes, chapitre 4). Ce résultat démontre que ce n'est pas la fourche en train de répliquer l'ADN qui maintient les contacts. De surcroît, ces contacts préférentiels persistent en dehors de la phase S, même pour des lymphocytes (cellules majoritaires des PBMC [[Delves et al., 2006](#)]) ayant effectué leur dernière phase S éventuellement plusieurs mois auparavant [[Sprent and Tough, 1994](#)]. Un corolaire de cette interprétation est le suivant : quelle que soit la manière dont l'organisation tri-dimensionnelle de la chromatine résulte du (ou conduit au) déroulement de la

réplication, celle-ci persiste pendant plusieurs mois en absence de contraintes liées à la réplication.

Le maintien d'associations à longue-distance en absence des processus nucléaires supposés les avoir établies a été observé plusieurs fois. [Palstra et al. \[2008\]](#) ont par exemple montré par 4C que les interactions à longue-distance du locus Rad23a (constitutivement exprimé) avec d'autres régions de chromatine active sont maintenues malgré l'inhibition de l'ARN-polymérase. Si l'activation de la transcription conduit à la formation de contacts à longue-distance avec d'autres régions actives [[Simonis et al., 2006](#)], ce n'est donc pas le fait d'être transcrit qui maintient les contacts. Cette affirmation est confirmée par l'étude de [Mitchell and Fraser \[2008\]](#) qui montre que les associations de gènes érythroïdes connus pour partager la même usine de transcription sont maintenues après l'inhibition de l'élongation. Il est également probable que ces associations soient également maintenues en absence d'initiation de la transcription, puisqu'elles persistent dans des cellules traitées à l' α -amanitine [[Palstra et al., 2008](#)].

Les contacts chromatinien à longue-distance, qu'ils soient observés entre régions co-exprimées ou entre régions co-répliquées, persistent donc en absence des polymérases actives.

La très grande similarité entre les profils 4C de PBMC dormantes et de PBMC en cycle (chapitre 4) est néanmoins surprenante. Ce résultat n'était pas réellement attendu car l'activation des PBMC par un traitement à la PHA est connue pour induire des modifications majeures, tant au niveau cytotologique [[Kyselá et al., 2005](#); [Branco and Pombo, 2006](#); [Branco et al., 2008](#)] qu'au niveau de l'expression génique [[Stentz and Kitabchi, 2004](#); [Argyropoulos et al., 2004](#); [Hess et al., 2004](#); [Chtanova et al., 2005](#); [Grigoryev et al., 2009](#)]. Malgré le doublement du volume nucléaire après la stimulation à la PHA [[Branco et al., 2008](#)], et malgré un niveau de compaction de la chromatine différent dans les lymphocytes dormants et activés au point d'empêcher certains anticorps d'atteindre leurs épitopes [[Baxter et al., 2004](#); [Chang and Aune, 2007](#); [Rawlings et al., 2011](#)], notre analyse 4C montre très peu de changements dans les partenaires à longue-distance. Ceci révèle une propriété surprenante des lymphocytes : les changements de l'agencement nucléaire visibles par cytologie se dérouleraient en conservant les contacts longue-distances (> 1 Mb) et en respectant donc un niveau d'organisation de la chromatine sous-jacent. Comme déjà discuté en fin du chapitre 4, une dilatation du noyau qui maintiendrait les contacts entre pics de timing ancrés par CTCF, autoriserait une décondensation globale de la chromatine (vue par cytologie) sans changer la ségrégation des séquences précoces et tardives dans l'espace nucléaire, et donc sans changer les relations de proximité relative. Cette interprétation serait au moins valable pour les 10 points de vue du chromosome 5 qui ont été les cibles de l'analyse 4C. Mais d'autres chromosomes pourraient avoir un comportement différent, puisque le chromosome 5 fait partie des chromosomes dont la taille absolue du territoire chromosomique change peu suite à l'activation par la PHA [[Branco et al., 2008](#)].

D'autres types cellulaires ne montrent pas la même structure nucléaire en phases G0 et G1. Le niveau de compaction de la chromatine est par exemple différent dans des fibroblastes en cycle et hors-cycle [[Bottiroli et al., 2004](#)], et la transition G0 $>$ G1 s'accompagne, dans ce type cellulaire, d'une modification de l'emplacement de certains territoires chromosomiques [[Bridger et al., 2000](#); [Mehta et al., 2010](#)]. L'organisation à grande échelle de la chromatine est également modifiée dans des hépatocytes entrant en cycle après une hépatectomie puisque l'ancrage à la matrice nucléaire est différent dans des hépatocytes en phase G0 (avant hépatectomie) et en cycle (24 h après hépatectomie) [[Rivera-Mulia et al., 2011](#)]. Il serait intéressant de tester si, comme pour les lymphocytes, le réagencement de la chro-

matine observé lorsque fibroblastes et hépatocytes entrent en cycle s'effectue en maintenant des contacts à longue-distance et en respectant donc une certaine organisation chromatinienne sous-jacente.

6.4- Forces pour établir cette organisation de la chromatine liée au timing de réplication

Une question encore en suspens concerne la nature des forces permettant l'établissement et le maintien de ces interactions. Nous avons déjà proposé que leur maintien puisse reposer, en tout cas en partie, sur CTCF qui est présent au niveau des maxima locaux de timing. Néanmoins, après chaque division cellulaire, cette organisation particulière de la chromatine doit se remettre en place, ce qui suppose, après la cytodièrese, que des mouvements de chromatine orientés existent, permettant la coalescence des régions précoces d'un côté, et des régions tardives de l'autre.

La chromatine peut diffuser dans le noyau à des vitesses qui devraient lui permettre d'explorer quelques μm en 1 h [Vazquez et al., 2001]. Ces mouvements pourraient favoriser la "rencontre" entre deux régions précoces après cytodièrese, autoriser leur interaction, laquelle serait ensuite maintenue, peut-être par CTCF. De manière intéressante, dans des cellules humaines, les vitesses de mouvement d'un locus unique [Thomson et al., 2004], mais aussi d'un territoire chromosomique entier [Walter et al., 2003], sont plus grandes en début de phase G1 que pendant le reste du cycle cellulaire. Ceci pourrait permettre une exploration plus grande du volume nucléaire lorsque l'organisation topologique de la chromatine doit se rétablir.

Des expériences récentes, certes dans le contexte de transcription, tendent à montrer que des mouvements de chromatine pourraient dépendre des moteurs de type actine-myosine. Chuang et al. [2006] ont analysé le repositionnement d'un locus, normalement près de l'enveloppe nucléaire, suite à son activation par un promoteur inductible. Le déplacement du locus comprend des phases de mouvements relativement stationnaires (où seule la diffusion brownienne intervient) interrompues par des phases de mouvements brusques probablement orientés (de l'ordre du μm en quelques minutes). De manière surprenante, l'expression de mutants de l'actine ou de la myosine nucléaire retarde, voire empêche, la relocalisation du locus, alors qu'elle est, au contraire, accélérée suite à l'expression d'un mutant stabilisant l'actine filamentueuse [Chuang et al., 2006]. L'implication de ces deux familles de protéines dans les mouvements de chromatine est confirmée par d'autres études [Dundr et al., 2007; Hu et al., 2008].

Si les mouvements de chromatine dépendent de l'actine et/ou de la myosine, connues pour permettre des mouvements orientés actifs, ils pourraient aussi être une conséquence passive d'un phénomène actif. En effet, l'actine et surtout les *actin-related proteins* (ARP) sont retrouvées dans de nombreux complexes protéiques capables de remodeler la chromatine [Oma and Harata, 2011]. Le recrutement d'un de ces complexes chez la levure, INO80, contenant différentes *actin-related proteins* (ARP4 –homologue de BAF53–, ARP5 ou ARP8), est capable d'augmenter la mobilité d'un locus dans le noyau [Neumann et al., 2012]. La fonction primaire de INO80 étant le remodelage de la chromatine à l'échelle du nucléosome [Oma and Harata, 2011], il est envisageable que, *via* le remodelage, les

propriétés physiques locales de la chromatine (longueur de persistance, densité) changent suffisamment la flexibilité naturelle de la fibre et modifie sa mobilité [Neumann et al., 2012]. Ainsi, le mouvement pourrait être une conséquence passive du remodelage actif de la chromatine.

Il manque encore des études montrant que ces acteurs sont aussi impliqués dans la mise en place de l'organisation de la chromatine en lien avec la réplication. Néanmoins, l'utilisation d'un anticorps monoclonal dirigé contre l'actine conduit, selon le protocole expérimentale utilisé (fixation méthanol, incubation de l'anticorps primaire toute la nuit à 4°C), à un marquage nucléaire [Fidlerová et al., 2005]. Les spots détectés dans ces conditions coïncident avec un marquage de l'ADN naissant de début de phase S. Bien que l'on ne puisse pas affirmer avec certitude que l'épitope détecté dans ces conditions expérimentales soit effectivement l'actine, cette étude suggère qu'une protéine possédant le motif SGP-SIVHRKCF (partie C-terminale de l'actine, utilisé comme immunogène) se localise aux foyers de réplication en début de phase S. D'autre part, une étude chez la levure montre que le complexe INO80, capable d'augmenter la cinétique de mouvement d'un locus (paragraphe précédent), est retrouvé au niveau des origines de réplication, notamment les plus précoces [Shimada et al., 2008]. La protéine ARP5, une de ses sous-unités, suit la même tendance, aussi bien en phase G1 que dans une population cellulaire asynchrone. Plusieurs études montrent en outre qu'INO80 est essentiel pour le déroulement de la réplication chez la levure [Shimada et al., 2008; Vincent et al., 2008; Papamichos-Chronakis and Peterson, 2008], mais aussi chez l'homme [Hur et al., 2010]. Il reste à démontrer que la capacité de se lier aux régions répliquées en début de phase S et de promouvoir les mouvements de chromatine sur de grandes distances sont transposables à l'homologue d'INO80 chez l'homme. Si c'est le cas, cela permettrait aux régions précoces de disposer d'un facteur enzymatique capable d'augmenter leur capacité à explorer un plus grand volume nucléaire, notamment en début de phase G1, lorsque doivent se reformer les contacts chromatinien entre loci précoce. Ceci pourrait alors favoriser leur rencontre, qui serait ensuite verrouillée par d'autres acteurs protéiques (CTCF).

6.5- Conclusion et perspectives

Le noyau est un organite très organisé. La chromatine qu'il contient est agencée de manière très spécifique, en une succession de repliements hiérarchiques très contrôlés. Si les premiers niveaux de repliement sont bien décrits, les niveaux supérieurs sont peu connus. Notamment, les déterminants qui les contrôlent, et les structures qui les stabilisent demeurent peu compris.

Mon travail de thèse a cherché, grâce à des approches 4C (Circular Chromosome Conformation Capture), à décrire au mieux l'organisation de la chromatine à grande échelle. Il a permis de montrer l'existence de liens étroits entre l'agencement tridimensionnel de la chromatine et l'ordre dans lequel d'ADN est dupliqué pendant la phase S (timing de réplication). Nous avons en effet démontré que deux séquences précoces consécutives, séparées par plus d'une Mb, s'associent dans le volume nucléaire, probablement par la formation de boucles englobant les séquences tardives intermédiaires. De plus, les régions tardives forment des domaines structuraux d'auto-interactions dont les frontières coïncident avec

les séquences les plus précoces. Enfin, pour de plus grandes séparations génomiques, la chromatine apparaît agencée de telle sorte que les séquences précoces et tardives occupent des territoires nucléaires distincts. Les profils de probabilité de contacts suivent alors admirablement le timing de réplication.

Alors que les liens entre agencement 3D de la chromatine et réplication semblent si étroits, des cellules en phase G0, dont la plupart avaient effectué leur dernière phase S des semaines auparavant, présentent une organisation topologique de la chromatine identique à celle capturée par 4C dans des cellules en cycle. Cela signifie que les contacts capturés ne résultent pas des ADN-polymérases en train de répliquer l'ADN.

La poursuite de ce travail descriptif passera nécessairement par l'éclaircissement des relations de cause et de conséquence (est-ce que la topologie de la chromatine reflète le déroulement de la réplication du cycle précédent, ou est-ce qu'elle contraint le déroulement de la phase S suivante?), ainsi que par la compréhension des forces permettant d'aboutir à cette organisation topologique. Ces questions seront, du reste, délicates à aborder puisqu'il est quasiment impossible de modifier l'organisation à grande échelle de la chromatine au niveau d'une région donnée. Par contre, il devrait être possible de tester si la structure de la chromatine reflète le déroulement de la réplication du cycle précédent. Pour cela, on pourrait influencer sur le déroulement de la réplication (par divers stress à l'hydroxyurée pour ralentir les fourches ou en utilisant le modèle expérimental de [Courbet et al. \[2008\]](#) dans lequel un mutant de cellules CHO peut être cultivé en condition où les fourches sont rapides ou lentes) et analyser la conformation de la chromatine par 4C au cycle suivant. Afin d'éviter tout artéfact induit par ces traitements chimiques, il serait intéressant d'appliquer la même stratégie sur des modèles de différenciation *in-vitro* (par exemple, la différenciation de cellules souches murines en précurseurs de crête neurale en une dizaine de jours) pour lesquels le timing de réplication est connu à différentes étapes. L'étude de la coordination des changements de timing de réplication avec ceux d'agencement 3D de la chromatine permettrait peut-être de préciser les relations de cause et de conséquence. Enfin, pour élucider le rôle de l'actine et/ou la myosine dans le réagencement de la chromatine après chaque division cellulaire, il pourrait être pertinent d'analyser les effets de l'inactivation des *actin-related proteins* (notamment celles retrouvées dans le complexe INO80), de la télophase au milieu de la phase G1. L'étude par 4C sera sans doute délicate, car elle nécessiterait d'analyser des cellules parfaitement synchrones, mais des mesures de proximité spatiale par FISH dans les cellules inactivées devraient être réalisables.

BIBLIOGRAPHIE

- Abdurashidova, G., Deganuto, M., Klima, R., Riva, S., Biamonti, G., Giacca, M., and Falaschi, A. (2000). Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science*, 287(5460):2023–2026.
- Adolph, K. W. (1980). Organization of chromosomes in HeLa cells: isolation of histone-depleted nuclei and nuclear scaffolds. *J Cell Sci*, 42:291–304.
- Adolph, K. W., Cheng, S. M., and Laemmli, U. K. (1977). Role of nonhistone proteins in metaphase chromosome structure. *Cell*, 12(3):805–816.
- Aladjem, M., Falaschi, A., and Kowalski, D. (2006). Chap.2 – Eukaryotic DNA Replication Origins. *Cold Spring Harbor Monograph Archive*, 47(0):31–61.
- Aladjem, M. I., Rodewald, L. W., Kolman, J. L., and Wahl, G. M. (1998). Genetic dissection of a mammalian replicator in the human beta-globin locus. *Science*, 281(5379):1005–1009.
- Allis, C. D., Berger, S. L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., Shilatifard, A., Workman, J., and Zhang, Y. (2007). New nomenclature for chromatin-modifying enzymes. *Cell*, 131(4):633–636.
- Anachkova, B., Djeliova, V., and Russev, G. (2005). Nuclear matrix support of DNA replication. *J Cell Biochem*, 96(5):951–961.
- Argyropoulos, C., Nikiforidis, G. C., Theodoropoulou, M., Adamopoulos, P., Boubali, S., Georgakopoulos, T. N., Paliogianni, F., Papavassiliou, A. G., and Mouzaki, A. (2004). Mining microarray data to identify transcription factors expressed in naïve resting but not activated T lymphocytes. *Genes Immun*, 5(1):16–25.
- Arneodo, A., Vaillant, C., Audit, B., Argoul, F., d’Aubenton Carafa, Y., and Thermes, C. (2011). Multi-scale coding of genomic information: From DNA sequence to genome structure and function. *Phys Rep*, 498:45–188.
- Audit, B., Zaghloul, L., Vaillant, C., Chevereau, G., d’Aubenton-Carafa, Y., Thermes, C., and Arneodo, A. (2009). Open chromatin encoded in DNA sequence is the signature of ‘master’ replication origins in human cells. *Nucleic Acids Res*, 37(18):6064–6075.
- Baù, D., Sanyal, A., Lajoie, B. R., Capriotti, E., Byron, M., Lawrence, J. B., Dekker, J., and Marti-Renom, M. A. (2011). The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol*, 18(1):107–114.
- Bacher, C. P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., and Heard, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol*, 8(3):293–299.
- Baddeley, D., Chagin, V. O., Schermelleh, L., Martin, S., Pombo, A., Carlton, P. M., Gahl, A., Domaing, P., Birk, U., Leonhardt, H., Cremer, C., and Cardoso, M. C. (2010). Measurement of replication structures at the nanometer scale using super-resolution light microscopy. *Nucleic Acids Res*, 38(2):e8.
- Baker, A., Audit, B., Chen, C.-L., Moindrot, B., Leleu, A., Guilbaud, G., Rappailles, A., Vaillant, C., Goldar, A., Mongelard, F., d’Aubenton-Carafa, Y., Hyrien, O., Thermes, C., and Arneodo, A. (2012). Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines. *PLoS Comput Biol*, 8:e1002443.
- Bantignies, F. and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. *Trends Genet*, 27(11):454–464.
- Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell*, 144(2):214–226.
- Baxter, J., Sauer, S., Peters, A., John, R., Williams, R., Caparros, M.-L., Arney, K., Otte, A., Jenuwein, T., Merkschlager, M., and Fisher, A. G. (2004). Histone hypomethylation is an indicator of epigenetic plasticity in quiescent lymphocytes. *EMBO J*, 23(22):4462–4472.
- Berezney, R. (2002). Regulating the mammalian genome: the role of nuclear architecture. *Adv Enzyme Regul*, 42:39–52.
- Berezney, R., Dubey, D. D., and Huberman, J. A. (2000). Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma*, 108(8):471–484.

- Bernardi, R. and Pandolfi, P. P. (2007). Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol*, 8(12):1006–1016.
- Biamonti, G., Perini, G., Weighardt, F., Riva, S., Giacca, M., Norio, P., Zentilin, L., Diviacco, S., Dimitrova, D., and Falaschi, A. (1992). A human DNA replication origin: localization and transcriptional characterization. *Chromosoma*, 102(1 Suppl):S24–S31.
- Bird, A. W., Yu, D. Y., Pray-Grant, M. G., Qiu, Q., Harmon, K. E., Megee, P. C., Grant, P. A., Smith, M. M., and Christman, M. F. (2002). Acetylation of histone H4 by Esa1 is required for DNA double-strand break repair. *Nature*, 419(6905):411–415.
- Bohn, M., Heermann, D. W., and van Driel, R. (2007). Random loop model for long polymers. *Phys Rev E*, 76(5 Pt 1):051805.
- Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Müller, S., Eils, R., Cremer, C., Speicher, M. R., and Cremer, T. (2005). Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol*, 3(5):e157.
- Bottiroli, G., Croce, A. C., Bottone, M. G., Vaccino, S., and Pellicciari, C. (2004). G0-G1 cell cycle phase transition as revealed by fluorescence resonance energy transfer: analysis of human fibroblast chromatin. *Eur J Histochem*, 48(1):37–48.
- Boyle, S., Gilchrist, S., Bridger, J. M., Mahy, N. L., Ellis, J. A., and Bickmore, W. A. (2001). The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet*, 10(3):211–219.
- Boyle, S., Rodesch, M. J., Halvensleben, H. A., Jeddelloh, J. A., and Bickmore, W. A. (2011). Fluorescence in situ hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Res*, 19(7):901–909.
- Branco, M. R., Branco, T., Ramirez, F., and Pombo, A. (2008). Changes in chromosome organization during PHA-activation of resting human lymphocytes measured by cryo-FISH. *Chromosome Res*, 16(3):413–426.
- Branco, M. R. and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol*, 4(5):e138.
- Bridger, J. M., Boyle, S., Kill, I. R., and Bickmore, W. A. (2000). Re-modelling of nuclear architecture in quiescent and senescent human fibroblasts. *Curr Biol*, 10(3):149–152.
- Brodie of Brodie, E. B., Nicolay, S., Touchon, M., Audit, B., d'Aubenton-Carafa, Y., Thermes, C., and Arneodo, A. (2005). From DNA sequence analysis to modeling replication in the human genome. *Phys Rev Lett*, 94(24):248103.
- Brown, J. M., Green, J., das Neves, R. P., Wallace, H. A. C., Smith, A. J. H., Hughes, J., Gray, N., Taylor, S., Wood, W. G., Higgs, D. R., Iborra, F. J., and Buckle, V. J. (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J Cell Biol*, 182(6):1083–1097.
- Brown, J. M., Leach, J., Reittie, J. E., Atzberger, A., Lee-Prudhoe, J., Wood, W. G., Higgs, D. R., Iborra, F. J., and Buckle, V. J. (2006). Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol*, 172(2):177–187.
- Bulger, M. and Groudine, M. (1999). Looping versus linking: toward a model for long-distance gene activation. *Genes Dev*, 13(19):2465–2477.
- Buongiorno-Nardelli, M., Micheli, G., Carri, M. T., and Marilley, M. (1982). A relationship between replicon size and supercoiled loop domains in the eukaryotic genome. *Nature*, 298(5869):100–102.
- Cadoret, J.-C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.-N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proc Natl Acad Sci U S A*, 105(41):15837–15842.
- Cai, S., Lee, C. C., and Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet*, 38(11):1278–1288.
- Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M. C., van Asperen, R., Boon, K., Voûte, P. A., Heisterkamp, S., van Kampen, A., and Versteeg, R. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science*, 291(5507):1289–1292.
- Carretero, M., Remeseiro, S., and Losada, A. (2010). Cohesin ties up the genome. *Curr Opin Cell Biol*, 22(6):781–787.
- Carter, D., Chakalova, L., Osborne, C. S., feng Dai, Y., and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nat Genet*, 32(4):623–626.

- Cayrou, C., Coulombe, P., Vigneron, A., Stanojic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., and Méchali, M. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Res*, 21(9):1438–1449.
- Chambeyron, S. and Bickmore, W. A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*, 18(10):1119–1130.
- Chambeyron, S., Silva, N. R. D., Lawson, K. A., and Bickmore, W. A. (2005). Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development*, 132(9):2215–2223.
- Chang, S. and Aune, T. M. (2007). Dynamic changes in histone-methylation 'marks' across the locus encoding interferon-gamma during the differentiation of T helper type 2 cells. *Nat Immunol*, 8(7):723–731.
- Chen, C.-L., Duquenne, L., Audit, B., Guilbaud, G., Rappailles, A., Baker, A., Huvet, M., d'Aubenton-Carafa, Y., Hyrien, O., Arneodo, A., and Thermes, C. (2011). Replication-associated mutational asymmetry in the human genome. *Mol Biol Evol*, 28(8):2327–2337.
- Chen, C.-L., Rappailles, A., Duquenne, L., Huvet, M., Guilbaud, G., Farinelli, L., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A., Hyrien, O., and Thermes, C. (2010). Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res*, 20(4):447–457.
- Chtanova, T., Newton, R., Liu, S. M., Weininger, L., Young, T. R., Silva, D. G., Bertoni, F., Rinaldi, A., Chappaz, S., Sallusto, F., Rolph, M. S., and Mackay, C. R. (2005). Identification of T cell-restricted genes, and signatures for different T cell responses, using a comprehensive collection of microarray datasets. *J Immunol*, 175(12):7837–7847.
- Chuang, C.-H., Carpenter, A. E., Fuchsova, B., Johnson, T., de Lanerolle, P., and Belmont, A. S. (2006). Long-range directional movement of an interphase chromosome site. *Curr Biol*, 16(8):825–831.
- Cockerill, P. N. and Garrard, W. T. (1986). Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell*, 44(2):273–282.
- Cook, P. R. (1999). The organization of replication and transcription. *Science*, 284(5421):1790–1795.
- Cook, P. R. (2002). Predicting three-dimensional genome structure from transcriptional activity. *Nat Genet*, 32(3):347–352.
- Cook, P. R., Brazell, I. A., and Jost, E. (1976). Characterization of nuclear structures containing superhelical DNA. *J Cell Sci*, 22(2):303–324.
- Courbet, S., Gay, S., Arnoult, N., Wronka, G., Anglana, M., Brison, O., and Debatisse, M. (2008). Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature*, 455(7212):557–560.
- Cremer, M., Küpper, K., Wagler, B., Wizelman, L., von Hase, J., Weiland, Y., Kreja, L., Diebold, J., Speicher, M. R., and Cremer, T. (2003). Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. *J Cell Biol*, 162(5):809–820.
- Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C., and Cremer, T. (2001). Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res*, 9(7):541–567.
- Cremer, T., Cremer, C., Baumann, H., Luedtke, E. K., Sperling, K., Teuber, V., and Zorn, C. (1982). Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Hum Genet*, 60(1):46–56.
- Cremer, T. and Cremer, M. (2010). Chromosome territories. *Cold Spring Harb Perspect Biol*, 2(3):a003889.
- Croft, J. A., Bridger, J. M., Boyle, S., Perry, P., Teague, P., and Bickmore, W. A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol*, 145(6):1119–1131.
- Cseresnyes, Z., Schwarz, U., and Green, C. M. (2009). Analysis of replication factories in human cells by super-resolution light microscopy. *BMC Cell Biol*, 10:88.
- Cuddapah, S., Jothi, R., Schones, D. E., Roh, T.-Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res*, 19(1):24–32.
- De, S. and Michor, F. (2011). DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. *Nat Biotechnol*, 29:1103–1108.
- de Laat, W. and Grosveld, F. (2003). Spatial organization of gene expression: the active chromatin hub. *Chromosome Res*, 11(5):447–459.
- de Wit, E. and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev*, 26(1):11–24.

- Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods*, 3(1):17–21.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science*, 295(5558):1306–1311.
- Delves, P. J., Martin, S. J., Burton, D. R., and Roitt, I. M. (2006). *Roitt's essential immunology*. Blackwell Pub., 11th edition edition.
- Depamphilis, M. and Blow, J. (2006). Chap.16 – Regulating Initiation Events in Metazoa. *Cold Spring Harbor Monograph Archive*, 47(0):313–334.
- Desprat, R., Thierry-Mieg, D., Lailier, N., Lajugie, J., Schildkraut, C., Thierry-Mieg, J., and Bouhassira, E. E. (2009). Predictable dynamic program of timing of DNA replication in human cells. *Genome Res*, 19(12):2288–2299.
- Dimitrova, D. S. and Gilbert, D. M. (1999). The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Mol Cell*, 4(6):983–993.
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485:376–380.
- Djeliova, V., Russev, G., and Anachkova, B. (2001a). Distribution of DNA replication origins between matrix-attached and loop DNA in mammalian cells. *J Cell Biochem*, 80(3):353–359.
- Djeliova, V., Russev, G., and Anachkova, B. (2001b). Dynamics of association of origins of DNA replication with the nuclear matrix during the cell cycle. *Nucleic Acids Res*, 29(15):3181–3187.
- Dobie, K. W., Lee, M., Fantes, J. A., Graham, E., Clark, A. J., Springbett, A., Lathe, R., and McClenaghan, M. (1996). Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. *Proc Natl Acad Sci U S A*, 93(13):6659–6664.
- Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., Green, R. D., and Dekker, J. (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 16(10):1299–1309.
- Drissen, R., Palstra, R.-J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev*, 18(20):2485–2490.
- Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., Shendure, J., Fields, S., Blau, C. A., and Noble, W. S. (2010). A three-dimensional model of the yeast genome. *Nature*, 465(7296):363–367.
- Dundr, M., Ospina, J. K., Sung, M.-H., John, S., Upender, M., Ried, T., Hager, G. L., and Matera, A. G. (2007). Actin-dependent intranuclear repositioning of an active gene locus in vivo. *J Cell Biol*, 179(6):1095–1103.
- Eaton, M. L., Galani, K., Kang, S., Bell, S. P., and MacAlpine, D. M. (2010). Conserved nucleosome positioning defines replication origins. *Genes Dev*, 24(8):748–753.
- Eaton, M. L., Prinz, J. A., MacAlpine, H. K., Tretyakov, G., Kharchenko, P. V., and MacAlpine, D. M. (2011). Chromatin signatures of the Drosophila replication program. *Genome Res*, 21(2):164–174.
- Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shores, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., and Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49.
- Farkash-Amar, S., Lipson, D., Polten, A., Goren, A., Helmstetter, C., Yakhini, Z., and Simon, I. (2008). Global organization of replication time zones of the mouse genome. *Genome Res*, 18(10):1562–1570.
- Farkash-Amar, S. and Simon, I. (2010). Genome-wide analysis of the replication program in mammals. *Chromosome Res*, 18(1):115–125.
- Fidlerová, H., Masata, M., Malinský, J., Fialová, M., Cvacková, Z., Louzecká, A., Koberna, K., Berezney, R., and Raska, I. (2005). Replication-coupled modulation of early replicating chromatin domains detected by anti-actin antibody. *J Cell Biochem*, 94(5):899–916.
- Filion, G. J., van Bommel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., Brugman, W., de Castro, I. J., Kerkhoven, R. M., Bussemaker, H. J., and van Steensel, B. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell*, 143(2):212–224.
- Finlan, L. E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J. R., and Bickmore, W. A. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet*, 4(3):e1000039.

- Fiorini, A., Gouveia, F., and Fernandez, M. A. (2006). Scaffold/Matrix Attachment Regions and intrinsic DNA curvature. *Biochemistry (Mosc)*, 71(5):481–488.
- Frank, A. C. and Lobry, J. R. (1999). Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene*, 238(1):65–77.
- Fraser, P. (2006). Transcriptional control thrown for a loop. *Curr Opin Genet Dev*, 16(5):490–495.
- Frouin, I., Montecucco, A., Spadari, S., and Maga, G. (2003). DNA replication: a complex matter. *EMBO Rep*, 4(7):666–670.
- Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H., Chew, E. G. Y., Huang, P. Y. H., Welboren, W.-J., Han, Y., Ooi, H. S., Ariyaratne, P. N., Vega, V. B., Luo, Y., Tan, P. Y., Choy, P. Y., Wansa, K. D. S. A., Zhao, B., Lim, K. S., Leow, S. C., Yow, J. S., Joseph, R., Li, H., Desai, K. V., Thomsen, J. S., Lee, Y. K., Karuturi, R. K. M., Herve, T., Bourque, G., Stunnenberg, H. G., Ruan, X., Cacheux-Rataboul, V., Sung, W.-K., Liu, E. T., Wei, C.-L., Cheung, E., and Ruan, Y. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, 462(7269):58–64.
- Gasser, S. M. and Laemmli, U. K. (1986). Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of *D. melanogaster*. *Cell*, 46(4):521–530.
- Gerdes, M. G., Carter, K. C., Moen, P. T., and Lawrence, J. B. (1994). Dynamic changes in the higher-level chromatin organization of specific sequences revealed by in situ hybridization to nuclear halos. *J Cell Biol*, 126(2):289–304.
- Gierman, H. J., Indemans, M. H. G., Koster, J., Goetze, S., Seppen, J., Geerts, D., van Driel, R., and Versteeg, R. (2007). Domain-wide regulation of gene expression in the human genome. *Genome Res*, 17(9):1286–1295.
- Gilbert, D. M., Takebayashi, S.-I., Ryba, T., Lu, J., Pope, B. D., Wilson, K. A., and Hiratani, I. (2010). Space and time in the nucleus: developmental control of replication timing and chromosome architecture. *Cold Spring Harb Symp Quant Biol*, 75:143–153.
- Girard-Reydet, C., Grégoire, D., Vassetzky, Y., and Méchali, M. (2004). DNA replication initiates at domains overlapping with nuclear matrix attachment regions in the xenopus and mouse c-myc promoter. *Gene*, 332:129–138.
- Goetze, S., Mateos-Langerak, J., Gierman, H. J., de Leeuw, W., Giromus, O., Indemans, M. H. G., Koster, J., Ondrej, V., Versteeg, R., and van Driel, R. (2007). The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol Cell Biol*, 27(12):4475–4487.
- Grigoryev, Y. A., Kurian, S. M., Nakorchevskiy, A. A., Burke, J. P., Campbell, D., Head, S. R., Deng, J., Kantor, A. B., Yates, J. R., and Salomon, D. R. (2009). Genome-wide analysis of immune activation in human T and B cells reveals distinct classes of alternatively spliced genes. *PLoS One*, 4(11):e7906.
- Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., Eussen, B. H., de Klein, A., Wessels, L., de Laat, W., and van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453(7197):948–951.
- Guilbaud, G., Rappailles, A., Baker, A., Chen, C.-L., Arneodo, A., Goldar, A., d'Aubenton Carafa, Y., Thermes, C., Audit, B., and Hyrien, O. (2011). Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS Comput Biol*, 7(12):e1002322.
- Habermann, F. A., Cremer, M., Walter, J., Kreth, G., von Hase, J., Bauer, K., Wienberg, J., Cremer, C., Cremer, T., and Solovei, I. (2001). Arrangements of macro- and microchromosomes in chicken cells. *Chromosome Res*, 9(7):569–584.
- Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., and Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc*, 2(7):1722–1733.
- Hakim, O., Sung, M.-H., Voss, T. C., Splinter, E., John, S., Sabo, P. J., Thurman, R. E., Stamatoyannopoulos, J. A., de Laat, W., and Hager, G. L. (2011). Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res*, 21(5):697–706.
- Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W. H., Ye, C., Ping, J. L. H., Mulawadi, F., Wong, E., Sheng, J., Zhang, Y., Poh, T., Chan, C. S., Kunarso, G., Shahab, A., Bourque, G., Cacheux-Rataboul, V., Sung, W.-K., Ruan, Y., and Wei, C.-L. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet*, 43(7):630–638.
- Hansen, R. S., Thomas, S., Sandstrom, R., Canfield, T. K., Thurman, R. E., Weaver, M., Dorschner, M. O., Gartler, S. M., and Stamatoyannopoulos, J. A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A*, 107(1):139–144.

- Hayashi, M., Katou, Y., Itoh, T., Tazumi, A., Tazumi, M., Yamada, Y., Takahashi, T., Nakagawa, T., Shirahige, K., and Masukata, H. (2007). Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *EMBO J*, 26(5):1327–1339.
- Heard, E. and Bickmore, W. (2007). The ins and outs of gene regulation and chromosome territory organisation. *Curr Opin Cell Biol*, 19(3):311–316.
- Hess, K., Yang, Y., Golech, S., Sharov, A., Becker, K. G., and Weng, N.-P. (2004). Kinetic assessment of general gene expression changes during human naive CD4+ T cell activation. *Int Immunol*, 16(12):1711–1721.
- Hewish, D. R. and Burgoyne, L. A. (1973). Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem Biophys Res Commun*, 52(2):504–510.
- Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazett-Jones, D. P., Plath, K., Dalton, S., Rathjen, P. D., and Gilbert, D. M. (2010). Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome Res*, 20(2):155–169.
- Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T. M., Schübeler, D., and Gilbert, D. M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol*, 6(10):e245.
- Horike, S., Cai, S., Miyano, M., Cheng, J.-F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet*, 37(1):31–40.
- Hou, C., Dale, R., and Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A*, 107(8):3651–3656.
- Hu, Q., Kwon, Y.-S., Nunez, E., Cardamone, M. D., Hutt, K. R., Ohgi, K. A., Garcia-Bassets, I., Rose, D. W., Glass, C. K., Rosenfeld, M. G., and Fu, X.-D. (2008). Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc Natl Acad Sci U S A*, 105(49):19199–19204.
- Hur, S.-K., Park, E.-J., Han, J.-E., Kim, Y.-A., Kim, J.-D., Kang, D., and Kwon, J. (2010). Roles of human INO80 chromatin remodeling enzyme in DNA replication and chromosome segregation suppress genome instability. *Cell Mol Life Sci*, 67(13):2283–2296.
- Huvet, M., Nicolay, S., Touchon, M., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A., and Thermes, C. (2007). Human gene organization driven by the coordination of replication and transcription. *Genome Res*, 17(9):1278–1285.
- Iborra, F. J., Pombo, A., Jackson, D. A., and Cook, P. R. (1996). Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *J Cell Sci*, 109 (Pt 6):1427–1436.
- Ikura, T., Ogryzko, V. V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell*, 102(4):463–473.
- Izzo, A. and Schneider, R. (2010). Chatting histone modifications in mammals. *Brief Funct Genomics*, 9(5-6):429–443.
- Jackson, D. A. and Pombo, A. (1998). Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. *J Cell Biol*, 140(6):1285–1295.
- Jiang, H. and Wong, W. H. (2008). SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics*, 24(20):2395–2396.
- Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J., and Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314):430–435.
- Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2012). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol*, 30(1):90–98.
- Karnani, N., Taylor, C. M., Malhotra, A., and Dutta, A. (2010). Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Mol Biol Cell*, 21(3):393–404.
- Keaton, M. A., Taylor, C. M., Layer, R. M., and Dutta, A. (2011). Nuclear scaffold attachment sites within ENCODE regions associate with actively transcribed genes. *PLoS One*, 6(3):e17912.
- Keys, J. R., Tallack, M. R., Zhan, Y., Papathanasiou, P., Goodnow, C. C., Gaensler, K. M., Crossley, M., Dekker, J., and Perkins, A. C. (2008). A mechanism for Ikaros regulation of human globin gene switching. *Br J Haematol*, 141(3):398–406.

- Kim, J., Guermah, M., McGinty, R. K., Lee, J.-S., Tang, Z., Milne, T. A., Shilatifard, A., Muir, T. W., and Roeder, R. G. (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell*, 137(3):459–471.
- Kitamura, E., Blow, J. J., and Tanaka, T. U. (2006). Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell*, 125(7):1297–1308.
- Kitsberg, D., Selig, S., Keshet, I., and Cedar, H. (1993). Replication structure of the human beta-globin gene domain. *Nature*, 366(6455):588–590.
- Knott, S. R. V., Peace, J. M., Ostrow, A. Z., Gan, Y., Rex, A. E., Viggiani, C. J., Tavaré, S., and Aparicio, O. M. (2012). Forkhead transcription factors establish origin timing and long-range clustering in *S. cerevisiae*. *Cell*, 148(1-2):99–111.
- Koberna, K., Ligasová, A., Malinský, J., Pliss, A., Siegel, A. J., Cvacková, Z., Fidlerová, H., Masata, M., Fialová, M., Raska, I., and Berezney, R. (2005). Electron microscopy of DNA replication in 3-D: evidence for similar-sized replication foci throughout S-phase. *J Cell Biochem*, 94(1):126–138.
- Koehler, D., Zakhartchenko, V., Froenicke, L., Stone, G., Stanyon, R., Wolf, E., Cremer, T., and Brero, A. (2009). Changes of higher order chromatin arrangements during major genome activation in bovine preimplantation embryos. *Exp Cell Res*, 315(12):2053–2063.
- Kohwi-Shigematsu, T. and Kohwi, Y. (1990). Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry*, 29(41):9551–9560.
- Koina, E. and Piper, A. (2005). An inactive X specific replication origin associated with a matrix attachment region in the human X linked HPRT gene. *J Cell Biochem*, 95(2):391–402.
- Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., and Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, 296(5565):158–162.
- Kouzarides, T. (2007a). Chromatin modifications and their function. *Cell*, 128(4):693–705.
- Kouzarides, T. (2007b). SnapShot: Histone-modifying enzymes. *Cell*, 131(4):822.
- Käs, E. and Chasin, L. A. (1987). Anchorage of the Chinese hamster dihydrofolate reductase gene to the nuclear scaffold occurs in an intragenic region. *J Mol Biol*, 198(4):677–692.
- Kumar, P. P., Bischof, O., Purbey, P. K., Notani, D., Urlaub, H., Dejean, A., and Galande, S. (2007). Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus. *Nat Cell Biol*, 9(1):45–56.
- Kumaran, R. I. and Spector, D. L. (2008). A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J Cell Biol*, 180(1):51–65.
- Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenko, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A*, 103(28):10684–10689.
- Kyselá, K., Philimonenko, A. A., Philimonenko, V. V., Janáček, J., Kahle, M., and Hozák, P. (2005). Nuclear distribution of actin and myosin I depends on transcriptional activity of the cell. *Histochem Cell Biol*, 124(5):347–358.
- Lagarkova, M. A., Svetlova, E., Giacca, M., Falaschi, A., and Razin, S. V. (1998). DNA loop anchorage region colocalizes with the replication origin located downstream to the human gene encoding lamin B2. *J Cell Biochem*, 69(1):13–18.
- Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat Cell Biol*, 9(10):1167–1174.
- Lemaitre, J.-M., Danis, E., Pasero, P., Vassetzky, Y., and Méchali, M. (2005). Mitotic remodeling of the replicon and chromosome structure. *Cell*, 123(5):787–801.
- León, P. and Kezer, J. (1990). Loop size in newt lampbrush chromosomes. *Chromosoma*, 99(1):83–86.
- Leonhardt, H., Rahn, H. P., Weinzierl, P., Sporbert, A., Cremer, T., Zink, D., and Cardoso, M. C. (2000). Dynamics of DNA replication factories in living cells. *J Cell Biol*, 149(2):271–280.
- Lewin, B. (2008). *Genes IX*. Jones & Bartlett Publishers Inc., U.S.; 9 edition, 9 edition edition.
- Li, B., Carey, M., and Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, 128(4):707–719.
- Li, F., Chen, J., Izumi, M., Butler, M. C., Keezer, S. M., and Gilbert, D. M. (2001). The replication timing program of the Chinese hamster beta-globin locus is established coincident with its repositioning near peripheral heterochromatin in early G1 phase. *J Cell Biol*, 154(2):283–292.

- Li, F., Chen, J., Solessio, E., and Gilbert, D. M. (2003). Spatial distribution and specification of mammalian replication origins during G1 phase. *J Cell Biol*, 161(2):257–266.
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., and Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–293.
- Liebich, I., Bode, J., Reuter, I., and Wingender, E. (2002). Evaluation of sequence motifs found in scaffold/matrix-attached regions (S/MARs). *Nucleic Acids Res*, 30(15):3433–3442.
- Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., Cherry, A. M., and Hoffman, A. R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science*, 312(5771):269–272.
- Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M. A., Clark, D., Kaur, M., Tandy, S., Kondoh, T., Rappaport, E., Spinner, N. B., Vega, H., Jackson, L. G., Shirahige, K., and Krantz, I. D. (2009). Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol*, 7(5):e1000119.
- Lobry, J. R. (1995). Properties of a general model of DNA evolution under no-strand-bias conditions. *J Mol Evol*, 40(3):326–330.
- Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol*, 13(5):660–665.
- Louwers, M., Splinter, E., van Driel, R., de Laat, W., and Stam, M. (2009). Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). *Nat Protoc*, 4(8):1216–1229.
- Lower, K. M., Hughes, J. R., Gobbi, M. D., Henderson, S., Viprakasit, V., Fisher, C., Goriely, A., Ayyub, H., Sloane-Stanley, J., Vernimmen, D., Langford, C., Garrick, D., Gibbons, R. J., and Higgs, D. R. (2009). Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. *Proc Natl Acad Sci U S A*, 106(51):21771–21776.
- Lubelsky, Y., Sasaki, T., Kuipers, M. A., Lucas, I., Beau, M. M. L., Carignon, S., Debatisse, M., Prinz, J. A., Dennis, J. H., and Gilbert, D. M. (2011). Pre-replication complex proteins assemble at regions of low nucleosome occupancy within the Chinese hamster dihydrofolate reductase initiation zone. *Nucleic Acids Res*, 39(8):3141–3155.
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648):251–260.
- Ma, H., Samarabandu, J., Devdhar, R. S., Acharya, R., Cheng, P. C., Meng, C., and Berezney, R. (1998). Spatial and temporal dynamics of DNA replication sites in mammalian cells. *J Cell Biol*, 143(6):1415–1425.
- MacAlpine, D. M., Rodríguez, H. K., and Bell, S. P. (2004). Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev*, 18(24):3094–3105.
- MacAlpine, H. K., Gordán, R., Powell, S. K., Hartemink, A. J., and MacAlpine, D. M. (2010). *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res*, 20(2):201–211.
- Mahy, N. L., Perry, P. E., and Bickmore, W. A. (2002). Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *J Cell Biol*, 159(5):753–763.
- Malyavantham, K. S., Bhattacharya, S., Alonso, W. D., Acharya, R., and Berezney, R. (2008). Spatio-temporal dynamics of replication and transcription sites in the mammalian cell nucleus. *Chromosoma*, 117(6):553–567.
- Mao, Y. S., Zhang, B., and Spector, D. L. (2011). Biogenesis and function of nuclear bodies. *Trends Genet*, 27(8):295–306.
- Markoulatos, P., Siafakas, N., and Moncany, M. (2002). Multiplex polymerase chain reaction: a practical approach. *J Clin Lab Anal*, 16(1):47–51.
- Marsden, M. P. and Laemmli, U. K. (1979). Metaphase chromosome structure: evidence for a radial loop model. *Cell*, 17(4):849–858.
- Martin, M. M., Ryan, M., Kim, R., Zakas, A. L., Fu, H., Lin, C. M., Reinhold, W. C., Davis, S. R., Bilke, S., Liu, H., Doroshov, J. H., Reimers, M. A., Valenzuela, M. S., Pommier, Y., Meltzer, P. S., and Aladjem, M. I. (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome Res*, 21(11):1822–1832.
- Mayer, R., Brero, A., von Hase, J., Schroeder, T., Cremer, T., and Dietzel, S. (2005). Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biol*, 6:44.
- McCready, S. J., Akrigg, A., and Cook, P. R. (1979). Electron-microscopy of intact nuclear dna from human cells. *J Cell Sci*, 39:53–62.

- Méchal, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol*, 11(10):728–738.
- Meaburn, K. J. and Misteli, T. (2007). Cell biology: chromosome territories. *Nature*, 445(7126):379–781.
- Mehta, I. S., Amira, M., Harvey, A. J., and Bridger, J. M. (2010). Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts. *Genome Biol*, 11(1):R5.
- Meister, P., Taddei, A., Ponti, A., Baldacci, G., and Gasser, S. M. (2007). Replication foci dynamics: replication patterns are modulated by S-phase checkpoint kinases in fission yeast. *EMBO J*, 26(5):1315–1326.
- Mesner, L. D., Crawford, E. L., and Hamlin, J. L. (2006). Isolating apparently pure libraries of replication origins from complex genomes. *Mol Cell*, 21(5):719–726.
- Mesner, L. D., Valsakumar, V., Karnani, N., Dutta, A., Hamlin, J. L., and Bekiranov, S. (2011). Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. *Genome Res*, 21(3):377–389.
- Mijalski, T., Harder, A., Halder, T., Kersten, M., Horsch, M., Strom, T. M., Liebscher, H. V., Lottspeich, F., de Angelis, M. H., and Beckers, J. (2005). Identification of coexpressed gene clusters in a comparative analysis of transcriptome and proteome in mouse tissues. *Proc Natl Acad Sci U S A*, 102(24):8621–8626.
- Miotto, B. and Struhl, K. (2010). HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. *Mol Cell*, 37(1):57–66.
- Mishiro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., Kinoshita, Y., and Nakao, M. (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J*, 28(9):1234–1245.
- Mitchell, J. A. and Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev*, 22(1):20–25.
- Münkel, C., Eils, R., Dietzel, S., Zink, D., Mehring, C., Wedemann, G., Cremer, T., and Langowski, J. (1999). Compartmentalization of interphase chromosomes observed in simulation and experiment. *J Mol Biol*, 285(3):1053–1065.
- Münkel, C. and Langowski, J. (1998). Chromosome structure predicted by a polymer model. *Phys. Rev. E*, 57(5):5888 – 5896.
- Moen, Jr, P. T., Johnson, C. V., Byron, M., Shopland, L. S., de la Serna, I. L., Imbalzano, A. N., and Lawrence, J. B. (2004). Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. *Mol Biol Cell*, 15(1):197–206.
- Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., and Duboule, D. (2011). A regulatory archipelago controls Hox genes transcription in digits. *Cell*, 147(5):1132–1145.
- Morey, C., Silva, N. R. D., Kmita, M., Duboule, D., and Bickmore, W. A. (2008). Ectopic nuclear reorganisation driven by a Hoxb1 transgene transposed into Hoxd. *J Cell Sci*, 121(Pt 5):571–577.
- Morey, C., Silva, N. R. D., Perry, P., and Bickmore, W. A. (2007). Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development*, 134(5):909–919.
- Mott, M. L. and Berger, J. M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Microbiol*, 5(5):343–354.
- Murrell, A., Heeson, S., and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat Genet*, 36(8):889–893.
- Nakayasu, H. and Berezney, R. (1989). Mapping replicational sites in the eucaryotic cell nucleus. *J Cell Biol*, 108(1):1–11.
- Nativio, R., Wendt, K. S., Ito, Y., Huddleston, J. E., Uribe-Lewis, S., Woodfine, K., Krueger, C., Reik, W., Peters, J.-M., and Murrell, A. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet*, 5(11):e1000739.
- Neumann, F. R., Dion, V., Gehlen, L. R., Tsai-Pflugfelder, M., Schmid, R., Taddei, A., and Gasser, S. M. (2012). Targeted ino80 enhances subnuclear chromatin movement and ectopic homologous recombination. *Genes Dev*, 26(4):369–383.
- Németh, A., Conesa, A., Santoyo-Lopez, J., Medina, I., Montaner, D., Péterfia, B., Solovei, I., Cremer, T., Dopazo, J., and Längst, G. (2010). Initial genomics of the human nucleolus. *PLoS Genet*, 6(3):e1000889.
- Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., Eussen, B., de Klein, A., Singer, R. H., and de Laat, W. (2011a). Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat Cell Biol*, 13(8):944–951.

- Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., Laat, W. D., and Duboule, D. (2011b). The dynamic architecture of Hox gene clusters. *Science*, 334(6053):222–225.
- Ohta, S., Tatsumi, Y., Fujita, M., Tsurimoto, T., and Obuse, C. (2003). The ORC1 cycle in human cells: II. Dynamic changes in the human ORC complex during the cell cycle. *J Biol Chem*, 278(42):41535–41540.
- O’Keefe, R. T., Henderson, S. C., and Spector, D. L. (1992). Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *J Cell Biol*, 116(5):1095–1110.
- Olins, A. L. and Olins, D. E. (1974). Spheroid chromatin units (v bodies). *Science*, 183(4122):330–332.
- Olins, D. E. and Olins, A. L. (2003). Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol*, 4(10):809–814.
- Oma, Y. and Harata, M. (2011). Actin-related proteins localized in the nucleus: from discovery to novel roles in nuclear organization. *Nucleus*, 2(1):38–46.
- Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W., and Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet*, 36(10):1065–1071.
- Osborne, C. S., Chakalova, L., Mitchell, J. A., Horton, A., Wood, A. L., Bolland, D. J., Corcoran, A. E., and Fraser, P. (2007). Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol*, 5(8):e192.
- Ottaviani, D., Lever, E., Takousis, P., and Sheer, D. (2008). Anchoring the genome. *Genome Biol*, 9(1):201.
- Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, 4(4):281–300.
- Paixão, S., Colaluca, I. N., Cubells, M., Peverali, F. A., Destro, A., Giadrossi, S., Giacca, M., Falaschi, A., Riva, S., and Biamonti, G. (2004). Modular structure of the human lamin B2 replicator. *Mol Cell Biol*, 24(7):2958–2967.
- Palstra, R.-J., Simonis, M., Klous, P., Brasset, E., Eijkelkamp, B., and de Laat, W. (2008). Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription. *PLoS One*, 3(2):e1661.
- Palstra, R.-J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet*, 35(2):190–194.
- Papamichos-Chronakis, M. and Peterson, C. L. (2008). The Ino80 chromatin-remodeling enzyme regulates replisome function and stability. *Nat Struct Mol Biol*, 15(4):338–345.
- Papapetrou, E. P., Ziros, P. G., Micheva, I. D., Zoumbos, N. C., and Athanassiadou, A. (2006). Gene transfer into human hematopoietic progenitor cells with an episomal vector carrying an S/MAR element. *Gene Ther*, 13(1):40–51.
- Parada, L. A., McQueen, P. G., and Misteli, T. (2004). Tissue-specific spatial organization of genomes. *Genome Biol*, 5(7):R44.
- Pardoll, D. M., Vogelstein, B., and Coffey, D. S. (1980). A fixed site of DNA replication in eucaryotic cells. *Cell*, 19(2):527–536.
- Pasero, P., Braguglia, D., and Gasser, S. M. (1997). ORC-dependent and origin-specific initiation of DNA replication at defined foci in isolated yeast nuclei. *Genes Dev*, 11(12):1504–1518.
- Patrinos, G. P., de Krom, M., de Boer, E., Langeveld, A., Imam, A. M. A., Strouboulis, J., de Laat, W., and Grosveld, F. G. (2004). Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev*, 18(12):1495–1509.
- Pauler, F. M., Sloane, M. A., Huang, R., Regha, K., Koerner, M. V., Tamir, I., Sommer, A., Aszodi, A., Jenuwein, T., and Barlow, D. P. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res*, 19(2):221–233.
- Paulson, J. R. and Laemmli, U. K. (1977). The structure of histone-depleted metaphase chromosomes. *Cell*, 12(3):817–828.
- Portela, A. and Esteller, M. (2010). Epigenetic modifications and human disease. *Nat Biotechnol*, 28(10):1057–1068.
- Postberg, J., Alexandrova, O., Cremer, T., and Lipps, H. J. (2005). Exploiting nuclear duality of ciliates to analyse topological requirements for DNA replication and transcription. *J Cell Sci*, 118(Pt 17):3973–3983.
- Postberg, J., Heyse, K., Cremer, M., Cremer, T., and Lipps, H. J. (2008). Spatial and temporal plasticity of chromatin during programmed DNA-reorganization in *Stylonychia* macronuclear development. *Epigenetics Chromatin*, 1(1):3.

- Prigent, C. and Dimitrov, S. (2003). Phosphorylation of serine 10 in histone H3, what for? *J Cell Sci*, 116(Pt 18):3677–3685.
- Qiu, X., Vu, T. H., Lu, Q., Ling, J. Q., Li, T., Hou, A., Wang, S. K., Chen, H. L., Hu, J. F., and Hoffman, A. R. (2008). A complex DNA looping configuration associated with the silencing of the maternal Igf2 allele. *Mol Endocrinol*, 22:1476–1488.
- Radichev, I., Parashkevova, A., and Anachkova, B. (2005). Initiation of DNA replication at a nuclear matrix-attached chromatin fraction. *J Cell Physiol*, 203(1):71–77.
- Raghuraman, M. K., Winzeler, E. A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D. J., Davis, R. W., Brewer, B. J., and Fangman, W. L. (2001). Replication dynamics of the yeast genome. *Science*, 294(5540):115–121.
- Rawlings, J. S., Gatzka, M., Thomas, P. G., and Ihle, J. N. (2011). Chromatin condensation via the condensin II complex is required for peripheral T-cell quiescence. *EMBO J*, 30(2):263–276.
- Razin, S. V. (2001). The nuclear matrix and chromosomal DNA loops: is there any correlation between partitioning of the genome into loops and functional domains? *Cell Mol Biol Lett*, 6(1):59–69.
- Razin, S. V., Mantieva, V. L., and Georgiev, G. P. (1979). The similarity of DNA sequences remaining bound to scaffold upon nuclease treatment of interphase nuclei and metaphase chromosomes. *Nucleic Acids Res*, 7(6):1713–1735.
- Reddy, K. L., Zullo, J. M., Bertolino, E., and Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, 452(7184):243–247.
- Remus, D., Beall, E. L., and Botchan, M. R. (2004). DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. *EMBO J*, 23(4):897–907.
- Rivera-Mulia, J. C., Hernández-Muñoz, R., Martínez, F., and Aranda-Anzaldo, A. (2011). DNA moves sequentially towards the nuclear matrix during DNA replication in vivo. *BMC Cell Biol*, 12:3.
- Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., Giraut, L., Després, B., Drevensek, S., Barneche, F., Dèrozier, S., Brunaud, V., Aubourg, S., Schnittger, A., Bowler, C., Martin-Magniette, M.-L., Robin, S., Caboche, M., and Colot, V. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J*, 30(10):1928–1938.
- Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T. C., Robins, A. J., Dalton, S., and Gilbert, D. M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res*, 20(6):761–770.
- Sachs, R. K., van den Engh, G., Trask, B., Yokota, H., and Hearst, J. E. (1995). A random-walk/giant-loop model for interphase chromosomes. *Proc Natl Acad Sci U S A*, 92(7):2710–2714.
- Sadoni, N., Cardoso, M. C., Stelzer, E. H. K., Leonhardt, H., and Zink, D. (2004). Stable chromosomal units determine the spatial and temporal organization of DNA replication. *J Cell Sci*, 117(Pt 22):5353–5365.
- Schaarschmidt, D., Baltin, J., Stehle, I. M., Lipps, H. J., and Knippers, R. (2004). An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *EMBO J*, 23(1):191–201.
- Schardin, M., Cremer, T., Hager, H. D., and Lang, M. (1985). Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Hum Genet*, 71(4):281–287.
- Schübeler, D., Scalzo, D., Kooperberg, C., van Steensel, B., Delrow, J., and Groudine, M. (2002). Genome-wide DNA replication profile for Drosophila melanogaster: a link between transcription and replication timing. *Nat Genet*, 32(3):438–442.
- Schepers, A. and Papior, P. (2010). Why are we where we are? Understanding replication origins and initiation sites in eukaryotes using ChIP-approaches. *Chromosome Res*, 18(1):63–77.
- Schoenfelder, S., Clay, I., and Fraser, P. (2010a). The transcriptional interactome: gene expression in 3D. *Curr Opin Genet Dev*, 20(2):127–133.
- Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J. A., Umlauf, D., Dimitrova, D. S., Eskiw, C. H., Luo, Y., Wei, C.-L., Ruan, Y., Bieker, J. J., and Fraser, P. (2010b). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*, 42(1):53–61.
- Sequeira-Mendes, J., Díaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gómez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. *PLoS Genet*, 5(4):e1000446.

- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell*, 148(3):458–472.
- Shimada, K., Oma, Y., Schleker, T., Kugou, K., Ohta, K., Harata, M., and Gasser, S. M. (2008). Ino80 chromatin remodeling complex promotes recovery of stalled replication forks. *Curr Biol*, 18(8):566–575.
- Shopland, L. S., Johnson, C. V., and Lawrence, J. B. (2002). Evidence that all SC-35 domains contain mRNAs and that transcripts can be structurally constrained within these domains. *J Struct Biol*, 140(1-3):131–139.
- Simonis, M., Klous, P., Homminga, I., Galjaard, R.-J., Rijkers, E.-J., Grosveld, F., Meijerink, J. P. P., and de Laat, W. (2009). High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat Methods*, 6(11):837–842.
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, 38(11):1348–1354.
- Simonis, M., Kooren, J., and de Laat, W. (2007). An evaluation of 3c-based methods to capture DNA interactions. *Nat Methods*, 4(11):895–901.
- Sparvoli, E., Levi, M., and Rossi, E. (1994). Replicon clusters may form structurally stable complexes of chromatin and chromosomes. *J Cell Sci*, 107 (Pt 11):3097–3103.
- Spector, D. L. (2006). SnapShot: Cellular bodies. *Cell*, 127(5):1071.
- Spector, D. L. and Lamond, A. I. (2011). Nuclear speckles. *Cold Spring Harb Perspect Biol*, 3(2):a000646.
- Spellman, P. T. and Rubin, G. M. (2002). Evidence for large domains of similarly expressed genes in the Drosophila genome. *J Biol*, 1(1):5.
- Spilianakis, C. G. and Flavell, R. A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol*, 5(10):1017–1027.
- Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R., and Flavell, R. A. (2005). Interchromosomal associations between alternatively expressed loci. *Nature*, 435(7042):637–645.
- Splinter, E., de Wit, E., Nora, E. P., Klous, P., van de Werken, H. J. G., Zhu, Y., Kaaij, L. J. T., van Ijcken, W., Gribnau, J., Heard, E., and de Laat, W. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev*, 25(13):1371–1383.
- Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*, 20(17):2349–2354.
- Sprent, J. and Tough, D. F. (1994). Lymphocyte life-span and memory. *Science*, 265(5177):1395–1400.
- Stentz, F. B. and Kitabchi, A. E. (2004). Transcriptome and proteome expression in activated human CD4 and CD8 T-lymphocytes. *Biochem Biophys Res Commun*, 324(2):692–696.
- Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765):41–45.
- Sueoka, N. (1995). Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. *J Mol Evol*, 40(3):318–325.
- Sun, Y., Durrin, L. K., and Krontiris, T. G. (2003). Specific interaction of PML bodies with the TP53 locus in Jurkat interphase nuclei. *Genomics*, 82(2):250–252.
- Sutherland, H. and Bickmore, W. A. (2009). Transcription factories: gene expression in unions? *Nat Rev Genet*, 10(7):457–466.
- Syed, S. H., Goutte-Gattat, D., Becker, N., Meyer, S., Shukla, M. S., Hayes, J. J., Everaers, R., Angelov, D., Bednar, J., and Dimitrov, S. (2010). Single-base resolution mapping of H1-nucleosome interactions and 3D organization of the nucleosome. *Proc Natl Acad Sci U S A*, 107(21):9620–9625.
- Szerlong, H. J. and Hansen, J. C. (2011). Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem Cell Biol*, 89(1):24–34.
- Tanabe, H., Müller, S., Neusser, M., von Hase, J., Calcagno, E., Cremer, M., Solovei, I., Cremer, C., and Cremer, T. (2002). Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A*, 99(7):4424–4429.
- Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J. R., Wickramasinghe, P., Lee, M., Fu, Z., and ichi Noma, K. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res*, 38(22):8164–8177.

- Tao, L., Dong, Z., Leffak, M., Zannis-Hadjopoulos, M., and Price, G. (2000). Major DNA replication initiation sites in the c-myc locus in human cells. *J Cell Biochem*, 78(3):442–457.
- Tark-Dame, M., van Driel, R., and Heermann, D. W. (2011). Chromatin folding—from biology to polymer models and back. *J Cell Sci*, 124(Pt 6):839–845.
- Thoma, F., Koller, T., and Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol*, 83(2 Pt 1):403–427.
- Thomson, I., Gilchrist, S., Bickmore, W. A., and Chubb, J. R. (2004). The radial positioning of chromatin is not inherited through mitosis but is established de novo in early G1. *Curr Biol*, 14(2):166–172.
- Thurman, R. E., Day, N., Noble, W. S., and Stamatoyannopoulos, J. A. (2007). Identification of higher-order functional domains in the human ENCODE regions. *Genome Res*, 17(6):917–927.
- Tolhuis, B., Blom, M., Kerkhoven, R. M., Pagie, L., Teunissen, H., Nieuwland, M., Simonis, M., de Laat, W., van Lohuizen, M., and van Steensel, B. (2011). Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet*, 7(3):e1001343.
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, 10(6):1453–1465.
- Touchon, M., Nicolay, S., Audit, B., Brodie of Brodie, E.-B., d'Aubenton-Carafa, Y., Arneodo, A., and Thermes, C. (2005). Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. *Proc Natl Acad Sci U S A*, 102(28):9836–9841.
- Truong, L. N. and Wu, X. (2011). Prevention of DNA re-replication in eukaryotic cells. *J Mol Cell Biol*, 3(1):13–22.
- Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., Weiss, M. J., Dekker, J., and Blobel, G. A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell*, 17(3):453–462.
- Valenzuela, M. S., Chen, Y., Davis, S., Yang, F., Walker, R. L., Bilke, S., Lueders, J., Martin, M. M., Aladjem, M. I., Massion, P. P., and Meltzer, P. S. (2011). Preferential localization of human origins of DNA replication at the 5'-ends of expressed genes and at evolutionarily conserved DNA sequences. *PLoS One*, 6(5):e17308.
- van den Engh, G., Sachs, R., and Trask, B. J. (1992). Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science*, 257(5075):1410–1412.
- Vassilev, L. and Johnson, E. M. (1990). An initiation zone of chromosomal DNA replication located upstream of the c-myc gene in proliferating HeLa cells. *Mol Cell Biol*, 10(9):4899–4904.
- Vaughn, J. P., Dijkwel, P. A., Mullenders, L. H., and Hamlin, J. L. (1990). Replication forks are associated with the nuclear matrix. *Nucleic Acids Res*, 18(8):1965–1969.
- Vazquez, J., Belmont, A. S., and Sedat, J. W. (2001). Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Curr Biol*, 11(16):1227–1239.
- Vernimmen, D., Marques-Kranc, F., Sharpe, J. A., Sloane-Stanley, J. A., Wood, W. G., Wallace, H. A. C., Smith, A. J. H., and Higgs, D. R. (2009). Chromosome looping at the human alpha-globin locus is mediated via the major upstream regulatory element (HS -40). *Blood*, 114(19):4253–4260.
- Versteeg, R., van Schaik, B. D. C., van Batenburg, M. F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H. J., and van Kampen, A. H. C. (2003). The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res*, 13(9):1998–2004.
- Vincent, J. A., Kwong, T. J., and Tsukiyama, T. (2008). ATP-dependent chromatin remodeling shapes the DNA replication landscape. *Nat Struct Mol Biol*, 15(5):477–484.
- Vogelstein, B., Pardoll, D. M., and Coffey, D. S. (1980). Supercoiled loops and eucaryotic DNA replicaton. *Cell*, 22(1 Pt 1):79–85.
- Walter, J., Schermelleh, L., Cremer, M., Tashiro, S., and Cremer, T. (2003). Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J Cell Biol*, 160(5):685–697.
- Waltz, S. E., Trivedi, A. A., and Leffak, M. (1996). DNA replication initiates non-randomly at multiple sites near the c-myc gene in HeLa cells. *Nucleic Acids Res*, 24(10):1887–1894.
- Wang, J., Shiels, C., Sasieni, P., Wu, P. J., Islam, S. A., Freemont, P. S., and Sheer, D. (2004a). Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic regions. *J Cell Biol*, 164(4):515–526.
- Wang, L., Lin, C.-M., Brooks, S., Cimbor, D., Groudine, M., and Aladjem, M. I. (2004b). The human beta-globin replication initiation region consists of two modular independent replicators. *Mol Cell Biol*, 24(8):3373–3386.

- Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A., and Feinberg, A. P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet*, 41(2):246–250.
- White, E. J., Emanuelsson, O., Scalzo, D., Royce, T., Kosak, S., Oakeley, E. J., Weissman, S., Gerstein, M., Groudine, M., Snyder, M., and Schübeler, D. (2004). DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. *Proc Natl Acad Sci U S A*, 101(51):17771–17776.
- Williams, A., Spilianakis, C. G., and Flavell, R. A. (2010). Interchromosomal association and gene regulation in trans. *Trends Genet*, 26(4):188–197.
- Williams, R. R. E., Broad, S., Sheer, D., and Ragoussis, J. (2002). Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Exp Cell Res*, 272(2):163–175.
- Woodfine, K., Beare, D. M., Ichimura, K., Debernardi, S., Mungall, A. J., Fiegler, H., Collins, V. P., Carter, N. P., and Dunham, I. (2005). Replication timing of human chromosome 6. *Cell Cycle*, 4(1):172–176.
- Woodfine, K., Fiegler, H., Beare, D. M., Collins, J. E., McCann, O. T., Young, B. D., Debernardi, S., Mott, R., Dunham, I., and Carter, N. P. (2004). Replication timing of the human genome. *Hum Mol Genet*, 13(2):191–202.
- Würtele, H. and Chartrand, P. (2006). Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res*, 14(5):477–495.
- Wu, C., Bassett, A., and Travers, A. (2007). A variable topology for the 30-nm chromatin fibre. *EMBO Rep*, 8(12):1129–1134.
- Wu, J. R. and Gilbert, D. M. (1996). A distinct G1 step required to specify the Chinese hamster DHFR replication origin. *Science*, 271(5253):1270–1272.
- Wyrick, J. J., Aparicio, J. G., Chen, T., Barnett, J. D., Jennings, E. G., Young, R. A., Bell, S. P., and Aparicio, O. M. (2001). Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science*, 294(5550):2357–2360.
- Xu, N., Tsai, C.-L., and Lee, J. T. (2006a). Transient homologous chromosome pairing marks the onset of X inactivation. *Science*, 311(5764):1149–1152.
- Xu, W., Aparicio, J. G., Aparicio, O. M., and Tavaré, S. (2006b). Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC Genomics*, 7:276.
- Yaffe, E., Farkash-Amar, S., Polten, A., Yakhini, Z., Tanay, A., and Simon, I. (2010). Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. *PLoS Genet*, 6(7):e1001011.
- Yaffe, E. and Tanay, A. (2011). Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet*, 43(11):1059–1065.
- Yokota, H., van den Engh, G., Hearst, J. E., Sachs, R. K., and Trask, B. J. (1995). Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *J Cell Biol*, 130(6):1239–1249.
- Yokoyama, T., Silversides, D. W., Waymire, K. G., Kwon, B. S., Takeuchi, T., and Overbeek, P. A. (1990). Conserved cysteine to serine mutation in tyrosinase is responsible for the classical albino mutation in laboratory mice. *Nucleic Acids Res*, 18(24):7293–7298.
- Yun, M., Wu, J., Workman, J. L., and Li, B. (2011). Readers of histone modifications. *Cell Res*, 21(4):564–578.
- Yusufzai, T. M. and Felsenfeld, G. (2004). The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proc Natl Acad Sci U S A*, 101(23):8620–8624.
- Zeit, M. J., Mukherjee, L., Bhattacharya, S., Xu, J., and Berezney, R. (2009). A probabilistic model for the arrangement of a subset of human chromosome territories in WI38 human fibroblasts. *J Cell Physiol*, 221(1):120–129.
- Zhang, Y., McCord, R. P., Ho, Y.-J., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., Becker, M. S., Alt, F. W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, 148(5):908–921.
- Zhao, R., Bodnar, M. S., and Spector, D. L. (2009). Nuclear neighborhoods and gene expression. *Curr Opin Genet Dev*, 19(2):172–179.
- Zhao, Z., Tavoosidana, G., Sjölander, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K. S., Singh, U., Pant, V., Tiwari, V., Kurukuti, S., and Ohlsson, R. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 38(11):1341–1347.

- Zink, D., Amaral, M. D., Englmann, A., Lang, S., Clarke, L. A., Rudolph, C., Alt, F., Luther, K., Braz, C., Sadoni, N., Rosenecker, J., and Schindelbauer, D. (2004). Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J Cell Biol*, 166(6):815–825.
- Zink, D., Cremer, T., Saffrich, R., Fischer, R., Trendelenburg, M. F., Ansorge, W., and Stelzer, E. H. (1998). Structure and dynamics of human interphase chromosome territories in vivo. *Hum Genet*, 102(2):241–251.
- Zorn, C., Cremer, C., Cremer, T., and Zimmer, J. (1979). Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. *Exp Cell Res*, 124(1):111–119.

Résumé :

L'organisation de la chromatine a une importance fonctionnelle pour contrôler le programme d'expression des gènes. Par contre, les liens qui l'unissent au déroulement de la réplication de l'ADN sont beaucoup moins connus.

Grâce à des approches basées sur la capture d'interactions chromosomiques et sur l'imagerie cellulaire, nous avons étudié les liens entre le repliement à grande échelle de la chromatine et le timing de réplication. Cette analyse, effectuée dans des cellules humaines lymphoblastoïdes, des cellules mononucléées du sang (PBMC) et des cellules issues d'une leucémie myéloïde à caractère érythrocytaire, a permis l'identification de domaines structuraux du noyau. Ces domaines sont relativement isolés les uns des autres et leurs frontières coïncident avec les zones d'initiation précoce. De plus, notre étude montre que ces zones d'initiation précoce interagissent préférentiellement, aussi bien entre voisins immédiats (séparation génomique de l'ordre de la mégabase) que le long du chromosome entier. Les loci répliqués tardivement interagissent eux-aussi avec leurs homologues, conduisant, dans l'espace nucléaire, à une ségrégation des loci en fonction de leur timing de réplication. Ces résultats sont soutenus par des mesures de distances sur des hybridations *in-situ* qui montrent que les loci répliqués en début de phase S sont plus proches qu'attendus.

Nos travaux révèlent enfin que l'organisation de la chromatine est similaire dans des cellules en phase G0 (PBMC dormantes), démontrant qu'elle n'est pas spécifique des cellules en phase S.

Pris ensemble, ces résultats apportent des preuves directes d'une organisation robuste de la chromatine, partagée par les cellules en cycle et dormantes, et corrélée au timing de réplication à différentes échelles.

Abstract :

Chromatin organization is of functional significance to control the gene expression program. However, its interplay with DNA replication program is less known.

Though the capture of chromosomal interactions and cell imaging, we studied the links between the high-order folding of chromatin and the replication timing. This analysis, which has been performed in human lymphoblastoid cells, in peripheral blood mononuclear cells (PBMC), and in a myeloid leukemia cell line with erythroid properties, allowed the identification of structural domains in the nucleus. These domains are quite isolated from each other and their boundaries coincide with early-initiation zones. In addition, our study shows that these early-initiation zones preferentially interact with each other. These interactions have been observed between neighboring early-initiation zones (genomic separation around 1 to few megabases) but also along the whole chromosome. The late-replicated loci interact with their counterparts as well, leading to a nuclear segregation of the loci according to their replication timing. These results are sustained by distance measurements in *in-situ* hybridizations which show that loci replicated at the beginning of S-phase are closer than expected.

Our work also reveals that chromatin organization is similar in cells blocked in G0 phase (quiescent PBMC), demonstrating that it does not result from the cells in S-phase.

Taken together, these results provide direct clues for a robust chromatin organization, common to cycling and resting cells, and related to the replication timing at different scales

Mot-clés : Chromatine, Réplication de l'ADN, Noyau cellulaire, Capture de conformation chromosomique, Génome

Keywords : Chromatin, DNA Replication, Cell nucleus, Chromosome Conformation Capture, Genome